

THESIS / THÈSE

MASTER EN SCIENCES INFORMATIQUES

Interaction vocale

Design de dialogue pour populations à besoins spécifiques

Walraet, Denis

Award date:
2019

Awarding institution:
Université de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



**UNIVERSITÉ
DE NAMUR**

FACULTÉ
D'INFORMATIQUE

**Interaction vocale : Design de dialogue pour
populations à besoins spécifiques**

Denys Walraet

À mon fils, Clément,
La plus belle des motivations.

Remerciements

Je tiens à remercier mon promoteur, Monsieur Bruno Dumas, Professeur à l'Université de Namur, pour son aide et ses précieux conseils dans la réalisation de ce mémoire. Son intérêt pour le sujet m'a motivé tout au long de ces mois de travail. J'ai apprécié ses avis critiques et ses corrections pour l'amélioration de la rédaction.

Mes remerciements vont également à tous mes proches, ma famille, mes amis et collègues pour leur compréhension et toutes leurs attentions.

Je remercie tout particulièrement ma compagne, Barbara, pour son soutien moral et ses encouragements qui m'ont permis de mener à terme ce projet. En espérant que les nombreuses concessions n'auront pas été vaines.

Résumé

Pour les populations à besoins spécifiques, l'interaction vocale est une solution idéale pour pallier les difficultés qu'ils rencontrent au quotidien. Les promesses d'accessibilité, de simplicité et d'efficacité en font une modalité de choix dans les environnements connectés et dans l'interaction avec les, désormais incontournables, assistants personnels. Cependant, les concepteurs ne bénéficient pas de règles, de principes, régissant la conception d'une interface vocale à destination de ces populations et ont bien souvent une méconnaissance des besoins qu'elles peuvent avoir. Il en résulte que bon nombre des solutions vocales présentes sur le marché négligent cette partie de la population.

Afin de répondre à cette problématique, ce mémoire présente un ensemble de trente recommandations visant à encadrer la conception d'une interface vocale à destination de populations à besoins spécifiques. Ces trente recommandations reposent sur une analyse des préoccupations et considérations des utilisateurs telles qu'elles sont mentionnées dans une série d'études sélectionnées pour leur pertinence face à la problématique. Dans le but d'aider les concepteurs dans leur tâche, treize heuristiques sont également introduites afin de fournir les outils nécessaires à une évaluation d'un système existant. Chaque recommandation se rapporte à une heuristique de sorte de disposer d'un ensemble de solutions pour une majorité des problèmes d'utilisabilité et, plus globalement d'expérience, rencontrés par les utilisateurs à besoins spécifiques.

Table des matières

1	Introduction	1
2	Interaction vocale et populations à besoins spécifiques : État des lieux	3
2.1	Une population vieillissante	3
2.2	Des besoins spécifiques	3
2.2.1	Amélioration de la qualité de vie	4
2.2.2	Maintien d'un lien social	4
2.2.3	Sentiment de sécurité	4
2.2.4	Suppression de la barrière technologique	4
2.3	La technologie comme solution envisageable	4
2.3.1	Solution omniprésente	5
2.3.2	Solution intelligente	5
2.3.3	Solution sécurisante	5
2.3.4	Solution vocale	6
2.4	Un fossé entre la technologie actuelle et les besoins spécifiques	6
2.4.1	Des problèmes d'utilisabilité	7
2.4.2	Des problèmes de fonctionnalité	7
2.4.3	Des problèmes de fiabilité	8
2.4.4	Des problèmes d'efficacité	8
2.4.5	Des problèmes de coûts	8
2.5	Une conception adaptée pour tous	9
2.5.1	Besoins spécifiques, handicaps, troubles divers et aptitudes	9
2.5.2	Universal design & ability-based design	10
2.5.3	User-centered design, participatory design & design for user empowerment	10
2.6	De la nécessité de recommandations de conception	11
3	Design de dialogue pour populations à besoins spécifiques : Évaluation à l'aide d'heuristiques	12
3.1	Définition des objectifs	12
3.2	Évaluation par heuristiques : Méthodologie	12
3.2.1	En quête de reconnaissance scientifique	13
3.2.2	Études de cas : Vue d'ensemble	14
3.2.3	Études de cas : Principales caractéristiques	17
3.2.4	Lien avec les heuristiques existantes	19
3.2.5	Définition des heuristiques	19
3.3	Évaluation par heuristiques : Résultats et interprétations	26
3.3.1	Contrôle sur le système	26
3.3.2	Visibilité du système et retour d'information	28
3.3.3	Prévention des erreurs	29
3.3.4	Reconnaissance des erreurs et mécanismes de correction	31

3.3.5	Stimulation de la découverte et de l'apprentissage	32
3.3.6	Aide et documentation	33
3.3.7	Modalités alternatives et efficacité	34
3.3.8	Conscience de l'utilisateur	35
3.3.9	Conscience du contexte d'utilisation	38
3.3.10	Respect de l'intimité de l'utilisateur et amélioration de sa sécurité	39
3.3.11	Entretien d'une relation de confiance basée sur la fiabilité et la transparence	40
3.3.12	Correspondance avec le monde réel	42
3.3.13	Support de l'intégration	42
3.4	Évaluation par heuristiques : Analyse et discussion	43
3.4.1	Validation des heuristiques	44
3.4.2	Comparaison entre les différentes populations	44
3.4.3	Préoccupations des populations à besoins spécifiques	46
4	Design de dialogue pour populations à besoins spécifiques : Élaboration des recommandations de conception	49
4.1	Principes de la conception basée sur les aptitudes	49
4.2	Recommandations pour les populations à besoins spécifiques	50
5	Conclusion	57
5.1	Limitations	57
5.1.1	Limitations des études sélectionnées	57
5.1.2	Limitations des recommandations de conception	58
5.2	Remarques finales et perspectives futures	58
	Références	59
A	Annexe : Préoccupations des populations à besoins spécifiques	65

Chapitre 1

Introduction

L'espérance de vie est en constante augmentation, nous sommes toujours plus nombreux à atteindre l'âge de la vieillesse et à vivre plus longtemps. Malheureusement, la vieillesse comporte son lot de désavantages et avec elle surgissent bien souvent des problèmes cognitifs, des troubles de la parole, une perte de l'acuité visuelle et des handicaps moteurs. Cependant, ces problèmes ne sont pas l'apanage de la vieillesse et touchent bon nombre de personnes à différents stades de leur vie.

Le fait de rencontrer des problèmes affectant la vie au quotidien induit des besoins spécifiques afin de pallier ces difficultés. Les besoins des populations touchées par un handicap consistent principalement en une amélioration de leur qualité de vie, un gain en autonomie et en indépendance. L'autonomie est définie comme étant la capacité à gérer les événements de sa vie et à prendre des décisions seul tandis que l'indépendance est le fait de se soustraire de l'assistance d'autrui pour accomplir une tâche.

Le recours à la technologie peut être envisagé pour répondre aux attentes des populations à besoins spécifiques. La technologie est omniprésente autour de nous. Elle peut être un outil promouvant l'accessibilité et la facilité de gestion des tâches quotidiennes. La technologie peut revêtir une forme d'intelligence propice à aider les utilisateurs dans la prise de décision et à comprendre leurs besoins. L'interaction vocale, en particulier, est une modalité naturelle, facile d'utilisation et opportune lorsqu'il s'agit de répondre à un manque d'accessibilité. L'usage de la parole pour interagir repose généralement sur des repères avec la réalité : les échanges conversationnels et la synthèse d'une voix naturelle. L'interaction vocale ne nécessite pas d'affinité particulière à la technologie ni de connaissances poussées en informatique. Elle favorise le contrôle de son environnement et la gestion des événements de la vie en centralisant l'interaction en un seul point. L'interaction vocale permet également de prendre en charge différents troubles limitant l'usage des modalités d'interaction traditionnelles. Cependant, la parole présente également des désavantages. C'est une modalité lente, éphémère, et qui demande de gros efforts cognitifs en matière d'apprentissage des commandes vocales et de compréhension du système. Elle induit souvent des attentes élevées de la part des utilisateurs liées à la ressemblance avec le monde réel. Enfin, elle manque parfois de fiabilité et peine à susciter son propre usage. Ces désavantages sont bien souvent atténués à l'aide d'un support multimodal. De fait, l'ajout d'une interface visuelle va souvent de pair avec l'interaction vocale.

Du point de vue des concepteurs d'un système intégrant une interface vocale, il est possible de pallier les désavantages de l'interaction grâce à des approches de conception orientées vers les utilisateurs finaux. Néanmoins, ces approches ne sont que des cadres orientant la conception sans pour autant fournir des éléments de solution à des problèmes donnés. Pour cette raison, des recommandations de conception sont nécessaires afin d'aider les concepteurs à répondre aux attentes des populations à besoins spécifiques.

Quelles sont les recommandations permettant de guider la conception d'un système intégrant

une interface vocale à destination de populations à besoins spécifiques? Pour répondre à cette problématique, les objectifs sont de déterminer un ensemble de problématiques rencontrées par les diverses populations et de les faire correspondre à des heuristiques. Ensuite de proposer pour chaque heuristiques des recommandations qui serviront à la conception d'un système disposant d'une interface vocale. Pour ce faire, il y a donc lieu de créer les heuristiques, de les valider en évaluant différents articles, de catégoriser l'ensemble des préoccupations des utilisateurs et de discuter des solutions envisageables.

Afin de remplir les objectifs répondant à la problématique posée, la méthodologie décrite ci-après a été appliquée :

1. Rechercher des documents pertinents à l'aide de critères excluant toute étude ne permettant pas d'établir les besoins de diverses populations, ciblées ou non. La volonté de ne pas se restreindre aux seules études relatives aux populations à besoins spécifiques est motivée par les travaux de Wobbrock *et al.*[81] sur la conception basée sur les aptitudes.
2. Selon la méthodologie de Rusu *et al.*[63], définir des heuristiques afin de catégoriser les besoins et préoccupations de ces populations. Pour ce faire, une étude exploratoire permet de mettre en lumière les caractéristiques générales de l'interaction et de les mettre en corrélation avec des heuristiques existantes, en l'occurrence celles de Nielsen[28] et d'Arhippainen[2]. En se basant sur les heuristiques existantes, créer de nouvelles heuristiques appliquées au domaine de l'interaction vocale et les décrire selon le template de Rusu *et al.*.
3. Analyser les articles sélectionnés afin d'identifier les principales préoccupations des utilisateurs ayant ou non des besoins spécifiques. Par la suite catégoriser les préoccupations en ayant recours aux heuristiques précédemment définies ainsi qu'interpréter les résultats ainsi obtenus.
4. Valider l'usage des nouvelles heuristiques pour évaluer les études relatives à l'interaction vocale. Confirmer ou infirmer la proposition de Wobbrock *et al.*[81], sur base des études présentant plusieurs dénominateurs communs, en déterminant s'il existe ou non des différences notables dans les préoccupations des populations à besoins spécifiques et des populations sans besoins spécifiques. En cas de réponse par l'affirmative, analyser minutieusement les seuls articles intégrant une évaluation par des utilisateurs à besoins spécifiques.
5. À l'aide de l'analyse des préoccupations des utilisateurs et des sept principes de la conception basée sur les aptitudes de Wobbrock *et al.*[82], rédiger des recommandations de conception pour l'interaction vocale entre un système et les populations sélectionnées pour chacune des heuristiques.

Ce travail est divisé en quatre parties. La première partie établit l'état des lieux de la technologie vocale en remplaçant l'interaction par la parole dans son contexte. La deuxième partie détermine les heuristiques utilisées pour catégoriser et comparer les besoins et préoccupations des différentes populations. La troisième partie définit les recommandations de conception sur base des préoccupations des populations à besoins spécifiques avant de conclure le travail, dans la dernière partie, avec les limitations connues et les perspectives futures.

Chapitre 2

Interaction vocale et populations à besoins spécifiques : État des lieux

2.1 Une population vieillissante

La population mondiale est vieillissante. C'est l'un des effets d'une espérance de vie en augmentation. Selon les derniers chiffres des Nations Unies[19], l'espérance de vie devrait mondialement augmenter pour atteindre 77 ans d'ici 2045-2050, au lieu de 71 ans à l'heure actuelle. Les projections prévoient que le nombre de personnes âgées de plus de 60 ans double d'ici la moitié de ce siècle pour atteindre 2.1 milliards. Sur la même période, c'est la tranche d'âge supérieure à 80 ans qui pourrait connaître la plus forte augmentation en passant de 137 millions à 425 millions d'individus.

Ce vieillissement n'est pas sans conséquence. Avec l'âge, les aptitudes cognitives et sensorielles diminuent[60] ainsi que les fonctions motrices. De nombreux handicaps peuvent survenir en vieillissant tels que des troubles mentaux, de la parole, de la vision, une perte de motricité ou même la surdité. Cependant ces handicaps ne sont malheureusement pas l'apanage de la vieillesse et touchent tout type de population.

Environ 10 pour cent de la population, soit 650 millions de personnes, vivent avec un handicap constituant ainsi la plus large minorité au monde. D'après l'Organisation mondiale de la santé (OMS), ce chiffre est en augmentation, suite à la croissance de la population, aux avancées médicales et au processus de vieillissement. Dans les pays où l'espérance de vie est de plus de 70 ans, chaque individu passera en moyenne huit ans ou 11,5 pour cent de sa vie à vivre avec un handicap[73]. Dans l'Union Européenne, des chiffres de 2012 indiquent qu'il y a plus de 73 millions de personnes de plus de 15 ans touchées par un handicap[22]. Certains de ces handicaps vont jusqu'à perturber la vie quotidienne de ces personnes. Ils sont ainsi près de 22 millions à déclarer avoir une difficulté dans leurs activités domestiques[21].

2.2 Des besoins spécifiques

Une population vieillissante et une large minorité affectée par un ou plusieurs handicaps, cela représente un groupe conséquent ayant des besoins spécifiques. Dans un tel groupe hétérogène, il existe des disparités entre les différents besoins exprimés. Ceux-ci varient en fonction du handicap et de sa sévérité. On peut néanmoins dégager des catégories de besoins principales ayant rencontrées le plus d'intérêt au sein de cette population telles que : l'amélioration de leur qualité de vie, leur sentiment de sécurité, le maintien d'un lien social et la suppression de la barrière technologique limitant leur accès à celle-ci.

2.2.1 Amélioration de la qualité de vie

L'amélioration de la qualité de vie est sans doute la catégorie regroupant les besoins les plus importants pour les personnes ayant un handicap. Un thème récurrent parmi les différentes études est le besoin d'indépendance. De nombreux actes du quotidien peuvent devenir de réelles corvées selon l'handicap concerné tels que l'interaction avec l'éclairage, le chauffage, les portes et fenêtres, les stores ou les appareils connectés fournissant divers services[17, 11, 68, 32]. En prenant le contrôle sur leur environnement, les personnes présentant divers troubles affectant l'accomplissement de leurs tâches domestiques se libèrent de la nécessité d'une assistance extérieure. Gagner en indépendance aussi longtemps et confortablement que possible concourt à leur bien-être mais agit également de manière positive sur leur sentiment de dignité[57] tout en satisfaisant leur souhait de limiter le recours à une aide institutionnelle en restant chez soi[68].

2.2.2 Maintien d'un lien social

Ne pas pouvoir sortir de chez soi ou avoir des difficultés à s'aventurer à l'extérieur du fait de problèmes moteurs contribue à isoler socialement cette partie de la population[54, 78]. L'un des désirs exprimés est de pouvoir rester en communication avec ses proches et le monde extérieur[11]. Ce désir est d'autant plus fort pour les individus plus jeunes pour lesquels les activités de divertissement et de socialisation sont plus importantes que la prise de contrôle sur l'environnement[32].

2.2.3 Sentiment de sécurité

Un environnement domestique n'est pas exempt de danger, d'autant plus lorsque l'on n'est pas en possession de toutes ses capacités motrices ou sensorielles. Le besoin de se sentir en sécurité est prévalent parmi la population la plus âgée, la littérature abondant en ce sens. Qu'il s'agisse d'un risque d'incendie, de fuite de gaz, de chute ou d'intrusion, leur souhait est d'être averti en cas de situation dangereuse mais également de notifier des personnes de leur entourage ainsi que les services d'aide appropriés[35, 57, 17]. Monitorer leurs fonctions vitales et éventuellement les partager avec leur docteur contribuerait également à les rassurer[17, 6].

2.2.4 Suppression de la barrière technologique

Améliorer la qualité de vie, maintenir le lien social et renforcer le sentiment de sécurité nécessite d'avoir recours à différents moyens techniques. Du fait de leur situation, une partie de la population touchée par un handicap éprouve des difficultés à interagir avec les outils technologiques. Leur souhait est de baisser cette barrière[36, 78] afin de profiter des avantages qu'a à leur offrir la technologie actuelle.

2.3 La technologie comme solution envisageable

Là où la plupart des personnes abordent les nouvelles technologies comme des gadgets divertissants[41, 59], la partie de la population à besoins spécifiques y voit des solutions afin de leur faciliter la vie, de gagner en confort et d'accroître leur sécurité[6, 32]. Il ne s'agit pas seulement de cette population mais également des gens qui les entourent, soignants ou membres de la famille. Ceux-ci apprécient le gain en indépendance de leur proche ou patient. Pour les familles, cela leur permet de les soulager de charges parfois trop lourdes en n'ayant plus à s'occuper d'actes du quotidien qui peuvent sembler anodins[68]. Pour les soignants professionnels, c'est le

transfert de leur disponibilité au profit des soins administrés aux patients[61]. C'est également un ensemble de moyens qui permet aux proches comme aux soignants de ne pas avoir peur de laisser ces personnes ayant divers troubles seules quelques instants.[61]

2.3.1 Solution omniprésente

Écrans aux murs ou jusqu'à être encastrés dans nos fours, frigos et machines-à-laver, enceintes intelligentes, capteurs en tous genres, ampoules connectées et verrous commandés à distance, la technologie est omniprésente autour de nous et presque indissociable de notre personne en nous accompagnant tout au long de la journée au poignet ou dans nos poches. Cette omniprésence telle qu'elle est présentée par Mark Weiser[79] se fond dans le décor au point qu'on ne la remarque plus, ou du moins, que l'on ne s'offusque pas de l'importance qu'elle ait dans notre quotidien. Du coup, comment ne pas être tenté de tirer profit de cette technologie toujours plus présente et toujours plus intelligente ? Les personnes à besoins spécifiques y voient une façon de combler une partie de leurs besoins[36]. Ils y voient une technologie qui même sous ses formes commerciales, et donc destinée au plus grand nombre, peut s'avérer accessible[58]. Un avantage certain des avancées techniques de ces dernières années est de rendre accessible de nombreuses technologies en les centralisant ou en les remplaçant en un seul point de contrôle. La centralisation favorise l'efficacité, l'indépendance et offre de nouvelles opportunités aux individus éprouvant des difficultés à utiliser la technologie en temps normal[58, 1]. Un sentiment également partagé par la population épargnée par le handicap[64, 8].

2.3.2 Solution intelligente

La technologie n'est pas seulement omniprésente, elle est également de plus en plus intrusive dans nos vies. Elle en apprend davantage sur nous à mesure que l'on interagisse avec et évolue en conséquence pour fournir des services en adéquation avec nos attentes. Enfin, elle est capable d'arborer un comportement quasi humain en proposant des échanges conversationnels[1]. La récolte des informations contextuelles, qu'elles soient relatives au bien-être du patient ou à son environnement, permet également d'avoir une meilleure appréhension de l'évolution de sa santé[44]. En prenant en compte le contexte dans lequel vivent les personnes à besoins spécifiques, la technologie est capable d'agir de manière pro-active, une qualité qui participe largement à son acceptation[11]. Au-delà de la conscience du contexte de l'utilisateur, c'est sa capacité à l'assister dans le déclin de ses fonctions cognitives en fournissant notamment un support mémoriel[58, 11] et un ensemble de routines évolutives à mesure que la maladie progresse[12].

2.3.3 Solution sécurisante

L'intelligence et l'omniprésence de la technologie permettent de mieux gérer les situations à risques. La multiplication de capteurs rend possible la détection de chutes[39] et facilite la prise en charge des appels de détresse même si la victime est loin d'un appareil de communication[75]. Les avancées technologiques ne se limitent pas à ces seuls cas de figure, les divers capteurs et systèmes existants contribuent à renforcer le sentiment de sécurité de la population y ayant recours aux dépens de leur intimité. Un inconvénient qui n'en est finalement pas un pour la population touchée par un handicap. Celle-ci semble peu concernée par les questions relatives à l'intrusion dans leur intimité et l'usage de leurs données personnelles[58, 36, 86]. Lorsque c'est le cas, elle en accepte volontiers les inconvénients tant les bénéfices qu'elle peut en tirer éclipsent les désavantages[17]. Un fait à mettre en relation avec la gravité du handicap : au plus le besoin d'avoir une aide se fait sentir, au plus la personne sera prête à faire des concessions sur

son intimité[6]. Certaines limites existent tout de même, l’usage de l’enregistrement vidéo n’est pas fort apprécié et ce surtout, de manière assez évidente, dans la salle de bain. Partager ces données avec des organismes gouvernementaux ou des compagnies d’assurance n’est également pas acceptable[6]. Seule exception, le partage avec son médecin si l’objectif est en lien avec une meilleure prise en charge de leur état de santé[80].

2.3.4 Solution vocale

Pour interagir avec tous ces systèmes, de nombreuses modalités existent. La coexistence de modalités, ou multimodalité, stimule plusieurs de nos sens simultanément et participe à la compréhension de l’échange avec le système. Pour les personnes à besoins spécifiques, elle facilite l’apprentissage de la technologie et leur permet de choisir la modalité la plus adéquate en fonction du contexte et de leurs aptitudes physiques ou cognitives. Un choix qui se fait de manière naturelle sans avoir à spécifier quelle interface utiliser au préalable[53]. La redondance de plusieurs moyens de communication est un élément important dans les problèmes d’interaction entre l’homme affecté par certaines incapacités et la machine[62].

La parole est un mode d’interaction majeur dans les systèmes multimodaux que ce soit en entrée ou en sortie[72]. La reconnaissance de la parole connaît un regain d’intérêt cette dernière décennie, grâce à la sortie commerciale de Siri[33], et gagne de plus en plus en maturité. C’est l’interface la plus naturelle et la plus facile à utiliser pour interagir avec la technologie[70, 67]. Elle ne nécessite pas de devoir traduire ses intentions ou d’apprendre des procédures complexes afin que le système exécute nos requêtes[69, 57]. Ordonner dans son propre langage avec une syntaxe et des mots-clés que l’on partage avec le système comme seules contraintes concourt largement à cette facilité d’utilisation. Sa nature omnidirectionnelle est un atout important comparé aux interfaces graphiques : il n’est plus nécessaire d’être en face d’un écran ou même d’avoir un retour visuel pour interagir avec le système[38] ce qui permet d’exécuter d’autres tâches en même temps[1, 14]. Un avantage qui en fait la modalité de prédilection des personnes ayant des troubles de la vision[38, 66, 4, 74, 1]. Les personnes âgées apprécient également d’utiliser la parole pour contrôler leur environnement[57]. Ils peuvent commander vocalement à leur rythme sans être dépassés par une technologie qui leur présente trop d’informations à la fois[36]. La population ayant des difficultés à se mouvoir ou à utiliser leurs membres supérieurs y trouvent une alternative séduisante aux modes d’interaction classiques[66, 25, 23, 75, 30]. En fonction de l’ampleur de leur handicap, la parole peut être utilisée conjointement avec d’autres modalités, le toucher principalement[53].

Bien qu’étant sur la voie de la maturité, la technologie vocale n’en reste pas moins imparfaite soit pour des raisons techniques comme la faiblesse du traitement automatique du langage naturel ou bien simplement d’un manque d’intelligence comme la compréhension profonde des intentions de l’utilisateur[69, 41]. Ce n’est pourtant pas un frein à l’adoption de la parole comme modalité de prédilection pour la majorité de la population à besoins spécifiques. En fait, son efficacité n’a que peu d’impact sur la satisfaction qu’en retirent ces personnes[85]. Ils acceptent volontiers les faiblesses du système[57] ce qui est à mettre en relation avec la sévérité de leur état[14] tout comme c’était le cas pour les problèmes soulevés quant au respect de leur intimité.

2.4 Un fossé entre la technologie actuelle et les besoins spécifiques

Même si l’interaction vocale présente de nombreux défauts et que ceux-ci sont globalement acceptés par les personnes n’ayant pas d’autre alternative pour contrôler les systèmes qui les entourent, il n’en reste pas moins qu’il n’est pas acceptable que ce soit eux qui s’adaptent à la

technologie et non l'inverse[32, 27]. Faire en sorte qu'elle soit aussi agréable pour ces personnes qu'elle ne peut l'être pour des individus sans handicap ne peut qu'améliorer leur qualité de vie et les rapports qu'ils ont à la technologie. De plus, certaines personnes parmi la population à besoins spécifiques éprouvent de grandes difficultés à interagir avec un système informatique en l'état actuel. Il est dès lors intéressant d'identifier les nombreux problèmes rencontrés par cette population afin de pouvoir au mieux y répondre. L'origine de ces problèmes n'est pas à chercher très loin. La majorité des avancées technologiques sont d'origines commerciales en ayant pour public cible des jeunes ne connaissant pas de troubles particuliers[57, 45, 71, 77]. Concevoir un système pour cette jeune population diffère d'une conception destinée à des populations plus fragiles. Non seulement leurs besoins divergent mais de plus les jeunes se montrent plus en mesure de s'adapter aux différentes faiblesses de conception[1].

2.4.1 Des problèmes d'utilisabilité

L'absence de support visuel augmente la charge cognitive de l'utilisateur lorsqu'il s'agit d'interagir exclusivement avec la parole[46, 3]. Un problème évident pour les personnes souffrants de troubles de la mémoire ayant du mal à se souvenir des instructions surtout sans le moindre rappel[58, 85]. Les problèmes d'apprentissage et de découverte des commandes vocales ne sont pas exclusifs à une population spécifique[58, 14], bon nombre d'études impliquant des utilisateurs sans handicap y font référence[24, 48, 26]. La nature éphémère de l'interaction vocale est à blâmer[66, 84] ainsi que son manque d'affordances et de métaphores. Ceci est également lié à un modèle mental défaillant des utilisateurs[66, 47] influencés notamment par leurs expériences passées. Enfin, un modèle mental erroné implique des attentes différentes de ce que le système est à même de fournir en matière d'intelligence, de capacités et d'objectifs[41, 13, 84].

Lorsqu'une erreur survient, il est compliqué de la corriger[66]. Un texte dicté contenant des fautes ou une commande non-reconnue demande du temps et de l'énergie pour y remédier. L'édition de texte par la parole est une opération fastidieuse pour les personnes malvoyantes, un véritable frein pour ceux qui n'utilisent pas la dictée de texte de manière régulière[4]. En cas de commande non-reconnue, il est difficile de savoir où se situe l'erreur : est-ce un problème de syntaxe ? de mot-clé ? d'audibilité ? Il en découle de la frustration[1, 85] et une impression de perdre le contrôle sur le système[42]. La population à besoins spécifiques veut être sûre de pouvoir interagir avec son environnement même si la reconnaissance de la parole venait à faillir et pour cela il lui faut pouvoir à tout moment revenir à des méthodes d'interaction plus traditionnelles[43, 87].

Plus insidieuse est la problématique de la dépendance sur la technologie vocale. Alors que celle-ci est supposée améliorer la qualité de vie des personnes ayant divers troubles et leur permettant de gagner en indépendance, c'est justement pour cette raison que ce groupe d'individus exprime ses craintes. Déléguer des tâches du quotidien au système par l'entremise de l'interaction vocale soulève des interrogations quant au déclin plus rapide de leurs capacités par manque d'activité[11] et de devenir complètement dépendant d'un système dont l'objectif était pourtant opposé[11, 57].

2.4.2 Des problèmes de fonctionnalité

La technologie vocale n'est pas adaptée à tous les handicaps. Les personnes atteintes de plusieurs handicaps, dont la surdité, n'y trouveront sûrement pas leur Graal dans l'interaction idéale avec un système. Pour ceux ayant des troubles de l'audition plus modérés, il est toujours possible d'augmenter le volume sonore du système[38]. Une possibilité qui n'est pas toujours disponible ou mal renseignée[58]. Autre population, autre problème, la population souffrant de dysarthrie ou de troubles limitant sa capacité à s'exprimer de manière continue pour laquelle

utiliser la parole comme mode d'interaction est une véritable épreuve. Les solutions commerciales actuelles manquent de considération pour ce groupe potentiel d'utilisateurs[58, 25] et offrent de faibles performances en cas d'utilisation par ceux-ci[5].

Même si la population à besoins spécifiques se montre indulgente en cas de non-respect de leur intimité par un système visant à améliorer leur qualité de vie, elle exprime son désaccord lorsqu'il est question d'utiliser une interface vocale en public. Elle ne désire pas que les personnes les entourant puissent entendre leurs échanges avec le système[4]. Outre ce problème relatif à leur vie privée, certains éprouvent une gêne d'entretenir un dialogue avec une interface vocale et regrettent que le système ne soit pas en mesure d'adapter l'interaction vocale en fonction du contexte[1]. Les systèmes basés sur l'interaction vocale manquent d'intelligence[66] et ne sont pas toujours en mesure de fournir des informations pertinentes à leurs utilisateurs[36].

Enfin, des problèmes de compatibilité sont soulevés du fait que la technologie vocale ne permet pas toujours de contrôler tous les appareils dans son environnement[36]. Différentes technologies cohabitent ensemble dans les environnements connectés et imposent des contraintes notamment structurelles à leurs utilisateurs[8].

2.4.3 Des problèmes de fiabilité

Dans un environnement dans lequel l'interaction vocale s'inscrit, il est nécessaire que la technologie propose un niveau de fiabilité satisfaisant pour que son usage soit adopté par un large public[8]. Les faiblesses du système sont généralement acceptées par la population à besoins spécifiques habituée à l'utiliser en l'état. Par contre, une minorité de cette population est plus critique envers les problèmes de fiabilité en n'envisage pas son adoption en cas d'erreurs[11]. Autre minorité, les personnes atteintes de dysarthrie pour lesquelles la technologie vocale ne rencontre pas leurs attentes[31]. Même si les problèmes de fiabilité sont globalement acceptés, le risque que le système devienne inopérant sans la moindre alternative pour prendre le relais est difficilement tolérable[57].

2.4.4 Des problèmes d'efficacité

L'engouement pour l'interaction vocale de ces dernières années est à mettre en corrélation avec les avancées commerciales des géants du Web (plus communément appelés les GAFA). Ceux-ci proposent depuis quelques années des appareils connectés sans écrans avec lesquels l'utilisateur interagit avec la parole et est informé en retour grâce à la synthèse vocale. Un usage qui pose nombre de problèmes[56]. Le retour d'informations sous forme vocale entre l'homme et la machine a les inconvénients majeurs d'être lent et temporel[66, 84] ce qui pose la question de l'efficacité d'une telle interface[1]. Suivant les usages, d'autres interfaces sont souvent conseillées car bien plus efficaces[23]. Non seulement la synthèse vocale ne permet pas de recevoir plusieurs informations à la fois, comme pourrait le faire une interface graphique[46], mais il est également difficile d'énoncer plusieurs commandes à la fois et d'espérer que le système les exécute toutes.

2.4.5 Des problèmes de coûts

Une dernière catégorie de problèmes est celle relative aux coûts d'un environnement contrôlable par la parole[58]. Bien que l'interaction vocale soit possible avec des appareils peu coûteux, son principal intérêt pour les populations à besoins spécifiques est d'améliorer leur qualité de vie. Il s'agit donc avant tout de leur permettre d'exécuter tout un ensemble de tâches quotidiennes en toute indépendance. Pour y parvenir, il leur faut un ensemble d'appareils contrôlables à distance tels que des ampoules, des verrous, des fenêtres, des stores, des thermostats, des appareils

connectés fournissant divers services, des moyens de communication, etc.[58, 11, 68, 32]. Pour ceux plus soucieux de leur sécurité, l'installation de plusieurs capteurs dans leur environnement sera également nécessaire[35]. Hormis le prix des composants, il faut également compter le prix de l'installation et ne pas négliger le coût des changements structurels nécessaires pour la mise en place du matériel. C'est ce dernier point qui est le plus problématique car installer ces appareils nécessite qu'ils soient raccordés au réseau électrique de l'environnement qu'il faudra également adapter[8]. Pour finir, il y a le coût de la maintenance à considérer.

2.5 Une conception adaptée pour tous

Comblant le fossé entre la technologie actuelle et les besoins d'une population qu'elle néglige demande de concevoir des systèmes prenant en compte les attentes de ces utilisateurs. Dans les domaines spécialisés, on parle de technologie d'assistance. Concevoir de tels systèmes pour une population à besoins spécifiques n'est pas chose aisée. Leurs pathologies sont très variées d'un individu à l'autre tout comme le sont leurs besoins[12, 18, 77]. Créer un système personnalisé pour chaque utilisateur, notamment pour la reconnaissance automatique du langage naturel, est une solution peu envisageable. Elle implique un investissement de l'utilisateur qui peut être fort contraignant pour les troubles les plus sévères[30] et des coûts supplémentaires rendant le projet financièrement non-viable[25]. En outre, ces troubles peuvent également évoluer avec le temps[60] nécessitant d'envisager des systèmes dynamiques capables de s'adapter à l'évolution des besoins des utilisateurs[85, 32, 18]. Pour composer avec cette diversité dynamique, la conception d'un système adaptatif permet d'en modifier l'interface utilisateur pour s'adapter à son usage, ses performances ou son état mental[58, 24, 64, 18].

2.5.1 Besoins spécifiques, handicaps, troubles divers et aptitudes

Avant d'envisager la réalisation d'un système accessible, il peut être nécessaire de revoir le concept de "population à besoins spécifiques". Dans la rédaction de ce mémoire, les termes "besoins spécifiques", "troubles" divers et "handicaps" sont utilisés sans distinction pour désigner les personnes ayant des déficiences physiques, mentales, intellectuelles ou sensorielles sur le long terme qui, en interaction avec divers obstacles, peuvent entraver leur participation pleine et effective à la société sur un pied d'égalité avec les autres[20]. Cependant, cette vision des choses repose sur des problèmes durables inhérents à leur personne auquel leur environnement n'est pas adapté. Selon Wobbrock[81], une autre manière d'aborder le sujet serait de considérer leurs déficiences comme un manque d'aptitudes à accomplir certaines tâches. Chaque individu a des aptitudes diverses lui permettant d'exécuter les actions de son quotidien mais il peut arriver qu'il n'y parvienne pas lorsque ces aptitudes viennent à manquer selon les circonstances. De ce fait, avec ou sans déficiences, tout le monde fait face à des situations dans lesquels il ne dispose pas des moyens nécessaires pour parvenir à ses fins. Les difficultés que nous rencontrons dans l'interaction avec notre environnement ne sont pas imputables à notre état mais à un manque de correspondance entre nos capacités et les hypothèses qui en sont faites par l'environnement. D'après Sears et Young[65], l'environnement de travail d'un individu et son contexte contribuent à l'existence d'handicaps, de déficiences et d'inaptitudes. Puisque nous éprouvons tous des difficultés, dans certaines circonstances, nous pouvons nous concentrer sur la diversité de nos aptitudes et non la diversité de nos déficiences. Concevoir un système selon nos aptitudes peut s'appliquer à chaque individu. Un exemple simple illustre parfaitement ce propos : une technologie conçue pour une personne n'ayant qu'un seul bras peut être tout autant utile pour une personne tenant son enfant dans l'un de ses bras[82].

2.5.2 Universal design & ability-based design

La notion d'aptitude et la manière dont elle peut influencer la réalisation d'un système se retrouve dans deux types de conceptions aux approches différentes : la conception universelle (ou universal design) et la conception basée sur les aptitudes (ou ability-based design[81]). La conception universelle tire son inspiration de l'architecture. Son objectif est de concevoir des structures, des environnements et par extension des produits pour que ceux-ci soient utilisables en prenant en compte le plus large éventail d'aptitudes possible. C'est une approche globale cherchant à satisfaire une grande diversité d'utilisateurs avec ou sans besoins spécifiques. Le désavantage de cette méthode est qu'elle propose une solution unique, universelle, pour y parvenir. Cela peut être une approche envisageable : il arrive qu'une bonne conception pour des personnes à besoins spécifiques soit une bonne conception pour le restant de la population également[85].

La conception basée sur les aptitudes met l'emphasis sur ce que les utilisateurs sont capables de faire et d'adapter le système ou l'environnement à eux et non l'inverse. Diverses stratégies existent comme l'adaptation automatique aux aptitudes, une grande configurabilité par l'utilisateur final, une personnalisation par une tierce partie ou en proposant diverses conceptions pour des aptitudes alternatives. Quelle que soit la stratégie choisie, ce sont les aptitudes de l'utilisateur qui contraignent le système à s'adapter ou à être adapté à celles-ci. Par opposition à la conception universelle, la conception basée sur les aptitudes a pour objectif de prendre en compte les aptitudes du plus grand nombre possible d'utilisateurs. La différence réside dans le fait de considérer individuellement les aptitudes de chacun plutôt que les aptitudes de l'ensemble. Cette approche permet la création d'interfaces générales ayant la flexibilité de s'adresser au plus grand nombre tout comme des interfaces sur-mesure pour des groupes plus restreints voire même des individus[82]. L'intelligence artificielle ainsi que les diverses technologies de capture sont des outils essentiels dans l'accomplissement de cet objectif.

2.5.3 User-centered design, participatory design & design for user empowerment

Les méthodes traditionnelles de développement proposent un cycle itératif reposant sur quatre piliers : l'analyse, la conception, le prototypage et enfin le test. Les designers et développeurs se basent sur ces quatre piliers pour concevoir des systèmes accessibles. Malheureusement ces piliers n'offrent pas suffisamment d'indications sur la conception d'un système adaptable. Ils se reposent donc sur des hypothèses erronées quant aux capacités des futurs utilisateurs. Ils tirent des leçons de leurs propres expériences, imaginent les comportements de leurs futurs utilisateurs et les représentent sous la forme d'un utilisateur unique pour lequel le système sera développé[49]. Le problème étant que lorsqu'il y a une dissonance entre les hypothèses des concepteurs et la réalité des utilisateurs c'est à ceux-ci que revient la contrainte de devoir s'adapter.

Quel que soit le public ciblé, il est nécessaire d'avoir recours à une conception centrée sur l'utilisateur[51]. Dans cette approche, les utilisateurs finaux sont inclus dans le processus de conception et de développement du système. De la sorte, on s'assure que la correspondance entre le système et les besoins des utilisateurs soit assurée, à minima, lors de la phase de test. Une implication si tard dans le processus ne permet pas de détecter les imperfections liées à l'analyse et la conception. Pour pallier ce problème, la conception participative propose d'impliquer plus tôt l'utilisateur final durant les phases de conception et de test. Un problème demeure, ces processus centrés sur l'utilisateur se focalisent sur des groupes homogènes d'utilisateurs représentés par des personnages fictifs, des personas, partageant les mêmes caractéristiques. Ils sont donc inefficaces pour rencontrer les besoins d'une population hétérogène ainsi que la

diversité dynamique des troubles affectant cette même population[85].

Une autre proposition de conception est celle de Richard Ladner[37]. La conception pour la responsabilisation de l'utilisateur, traduction littérale de "design for user empowerment", va plus loin dans l'implication de l'utilisateur final au point que celui-ci est présent dans chacune des quatre étapes du cycle de conception. Ce sont les utilisateurs qui prennent en charge le projet, déterminent les exigences et les fonctionnalités, développent les prototypes, les testent, et itèrent sur les résultats pour raffiner la conception. Dans le cas d'un système à destination de personnes à besoins spécifiques, ce sont des individus différents, techniquement compétents, et partageant les mêmes besoins qui ont la responsabilité de chacune des étapes. Malheureusement, les compétences techniques nécessaires ne se retrouvent pas toujours au sein de la population ciblée et certains troubles affectant les capacités cognitives peuvent d'autant plus compliquer l'accès à cette expertise technique. Toujours est-il que l'implication des utilisateurs finaux au projet est la meilleure assurance de parvenir à un résultat répondant à l'ensemble de leurs attentes.

2.6 De la nécessité de recommandations de conception

Il existe plusieurs approches et méthodes de conception permettant de rejoindre les besoins spécifiques de diverses populations. Elles n'impliquent pas pour autant que la facilité d'utilisation du produit final soit respectée. Selon l'Organisation internationale de normalisation, l'utilisabilité est définie comme étant le degré selon lequel un système, un produit ou un service peut être utilisé par des utilisateurs spécifiques pour accomplir des objectifs définis avec efficacité, efficience et satisfaction dans un contexte d'utilisation précis[9]. Les méthodes d'inspection permettent de trouver les problèmes de facilité d'utilisation en se basant sur l'expertise de professionnels. L'évaluation de l'utilisabilité à l'aide d'heuristiques est l'une de ces méthodes.

Il y a plus de 25 ans, Nielsen a rédigé dix heuristiques[28] pensées comme des règles générales s'appliquant à la facilité d'utilisation des interfaces graphiques. Cette méthode d'évaluation est facile à exécuter, peu coûteuse et permet de détecter tout un ensemble de problèmes relatifs à l'interaction entre l'homme et la machine. Cependant, elle n'est pas toujours adaptée aux problèmes spécifiques à un domaine particulier. Il y a donc un besoin de définir de nouvelles heuristiques et spécialement pour les nouveaux systèmes d'information apportant leur lot de challenges[63]. Dans le domaine de l'interaction vocale, Murad *et al.* ont jeté les bases d'un tel travail en proposant de vérifier la pertinence de l'usage des dix heuristiques pour catégoriser les problèmes d'utilisabilité discutés dans une vingtaine d'études[46]. Leur travail tend à prouver que, moyennant quelques adaptations et une actualisation au contexte actuel, ces règles générales englobent la majorité des problèmes rencontrés.

S'assurer de la facilité d'utilisation d'un système ne garantit pas le plaisir que peut en retirer l'utilisateur final ni que ses besoins soient satisfaits. L'utilisabilité fait partie d'un tout, l'expérience utilisateur, qui comprend tous les aspects de l'interaction entre les utilisateurs et le système[29]. À la manière de Nielsen, Arhipainen a rédigé 10 heuristiques applicables à l'expérience utilisateur[2]. Ses heuristiques sont générales et non-spécifiques à un domaine particulier nécessitant quelques adaptations.

Chapitre 3

Design de dialogue pour populations à besoins spécifiques : Évaluation à l'aide d'heuristiques

3.1 Définition des objectifs

Ce mémoire vise à faciliter l'utilisation des interfaces vocales par les populations à besoins spécifiques, à améliorer leur expérience d'une telle technologie, et à promouvoir l'usage de la parole en tant que principale modalité d'interaction pour l'accomplissement de tâches domestiques. Pour ce faire, différents objectifs sont définis :

- déterminer les principales problématiques rencontrées par les diverses populations en se basant sur des études de cas incluant une expérimentation avec des utilisateurs,
- définir un ensemble d'heuristiques applicables aux problématiques précédemment listées,
- évaluer les différentes heuristiques définies en leur faisant correspondre les problématiques,
- établir des recommandations pour chacune des heuristiques qui serviront à la conception d'un système intégrant une interface vocale comme principale modalité.

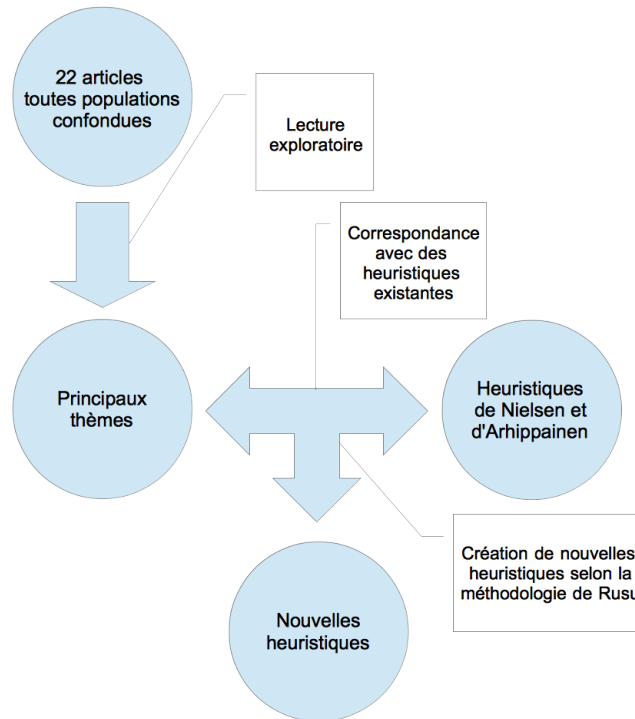
La liste des différentes études, les principales problématiques mentionnées dans les articles, et la définition de nouvelles heuristiques, sont présentées dans la section deux. La section trois aborde l'ensemble des problématiques décelées en les regroupant par heuristique correspondante. La section quatre propose une validation des heuristiques, une comparaison entre les préoccupations des diverses populations, et une discussion sur les préoccupations des populations à besoins spécifiques.

3.2 Évaluation par heuristiques : Méthodologie

Résumé

Afin de mener à terme les objectifs de ce mémoire, vingt-deux articles ont été sélectionnés conformément à une série de critères. Plusieurs populations, à besoins spécifiques ou non, sont représentées en se basant sur le travail de Wobbrock et al.[82]. Selon eux, les différentes populations rencontrent les mêmes besoins en fonction des situations, et ce, indépendamment de leur état physique ou mental. Une analyse des études permet de mettre en évidence les préoccupations des utilisateurs qui serviront à l'élaboration des recommandations de conception. Afin de catégoriser les diverses préoccupations, un ensemble d'heuristiques sont créées en appliquant la

méthodologie de Rusu et al.[63]. Suite à une lecture exploratoire, les principaux thèmes abordés par les utilisateurs relatifs à l'utilisabilité et leur expérience du système sont dégagés des vingt-deux articles. Ces thèmes sont mis en relation avec des heuristiques existantes, en l'occurrence celles de Nielsen[28] et d'Arhippainen[2]. Enfin, treize nouvelles heuristiques sont définies sur base des heuristiques existantes.



Graphique 3.1 – Création des heuristiques

3.2.1 En quête de reconnaissance scientifique

Définir un ensemble de recommandations nécessite une compréhension des problématiques relatives à l'interaction vocale pour des populations à besoins spécifiques. Une analyse de plusieurs cas d'étude permet de mettre en lumière ces problématiques et les éventuelles solutions proposées. Une recherche a été réalisée avec quelques mots-clés spécifiques pour dégager une série de documents, à caractère scientifique, dignes d'intérêt (voir tableau 3.1). En itérant systématiquement sur les différentes références des documents, le champs de recherche s'est considérablement élargi. Les articles antérieurs à l'avènement des principaux assistants vocaux tels que Siri, Cortana, Alexa et Google Assistant ont été sélectionnés avec précaution. La raison derrière ce choix est que ces différentes technologies ont redéfini l'interaction vocale entre l'homme et la machine influençant par la même occasion les modèles mentaux et les attentes des utilisateurs. De plus, les cas d'étude exempts d'évaluation par de futurs utilisateurs, ou de recommandations de conception à leur intention, n'ont pas été retenus pour une analyse ultérieure. Enfin, l'interaction vocale faisant partie intégrante des environnements connectés, quelques articles relatifs à ce domaine sont également inclus.

Bien qu'il existe une kyrielle d'études ayant un lien avec l'interaction vocale, très peu d'entre elles fournissent des éléments pertinents à l'élaboration de recommandations de conception, toutes populations confondues. Nombre de ces études ont des considérations techniques visant à améliorer la reconnaissance automatique de la parole ou n'incluent pas d'expérimentation avec des utilisateurs. Au final, seuls vingt-deux articles conformes aux critères de recherche ont été retenus pour faire l'objet d'une analyse. Le faible nombre d'études sélectionnées illustre un

certain manque d'intérêt de la part de la communauté scientifique pour l'amélioration de l'interaction vocale en matière d'utilisabilité et d'expérience utilisateur. L'interaction vocale n'est pas encore considérée comme une modalité de premier plan, la faute aux inconvénients intrinsèques de cette technologie[66] et à son usage principalement ludique[59, 41]. À titre de comparaison, six ans se sont écoulés entre l'avènement de l'interface graphique sur les ordinateurs destinés au grand public¹ et la proposition des dix heuristiques de Nielsen et Molich[50]. Siri a marqué une nouvelle ère dans l'interaction vocale depuis huit ans déjà[52]² et un manque de principes et de recommandations de conception se fait sentir. Murad *et al.* ont proposé une première approche en appliquant les heuristiques de Nielsen à différentes études relatives à l'interaction vocale[46]. Leurs conclusions invitent les chercheurs en interaction homme-machine à réaliser le même travail que Nielsen et Molich ont pu réaliser il y a près de trois décennies.

Contrairement au travail de Murad *et al.*, la grande majorité des articles sélectionnés sont des cas d'étude appuyés par une expérimentation avec des utilisateurs ayant ou non des besoins spécifiques. Ce choix d'inclure les populations à besoins spécifiques aux autres populations est motivé par l'hypothèse que, quelque soit la population ciblée, les problématiques rencontrées sont les mêmes. Seules certaines recommandations diffèrent, reposant sur les aptitudes des utilisateurs pour proposer des solutions alternatives. Cette hypothèse s'appuie sur le travail de Wobbrock *et al.*[81] invitant à concevoir des systèmes en considérant les aptitudes des utilisateurs et non leur manque d'aptitude³.

voice speech	interface interaction recognition assistant control hci	learnability usability security satisfaction evaluation design guidelines	impairment disabilities paralyzed - paralysis paraplegia mobility impaired reduced physical ability dysarthric blind - low vision deaf	home automation home appliances smart home assistive technology
-----------------	--	---	--	--

Tableau 3.1 – Liste des mots-clés utilisés

3.2.2 Études de cas : Vue d'ensemble

1. Macintosh 128k d'Apple en 1984

2. Lancement de Siri d'Apple en 2011

3. Selon une série de facteurs situationnels identifiés par Wobbrock *et al.*, les différentes populations rencontrent des situations où leurs besoins sont identiques. Cela se traduit par des limitations dans leurs aptitudes physiques et cognitives affectant l'interaction des utilisateurs avec la technologie. Ils en concluent qu'une conception orientée sur les aptitudes conduit à des systèmes plus accessibles et utilisables pour tout le monde.

Ref.	Auteur(année)	Titre	Objectif	Groupe ciblé	Modalité d'entrée	Modalité de sortie	Style d'interaction	Contexte d'utilisation
[58]	Pradhan <i>et al.</i> (2018)	« Accessibility came by accident » : use of voice-controlled intelligent personal assistants by people with disabilities	Étudier le degré d'accessibilité des assistants personnels intelligents (Amazon Echo & Dot, Google Home)	Personnes avec des déficiences et plus précisément les personnes malvoyantes	Parole	Son	Conversationalnel	Assistant intelligent (enceinte connectée)
[57]	Portet <i>et al.</i> (2013)	Design and evaluation of a smart home voice interface for the elderly : acceptability and objection aspects	Évaluer l'acceptation de l'environnement connecté Sweet-Home par des personnes âgées autonomes	Personnes âgées	Parole	Son (appels et feedback du système seulement), affichage sur écran	Commande vocale	Habitat intelligent, domotique
[14]	Corbett <i>et al.</i> (2016)	What can I say ? Addressing user experience challenges of a mobile voice user interface for accessibility	Améliorer la visibilité et l'apprentissage des commandes vocales d'une application mobile avec une interface vocale (projet VoiceNavigator)	Personnes avec troubles moteurs affectant leur dextérité des mains	Parole	Son (appels seulement), affichage sur écran	Commande vocale	Application mobile
[75]	Vacher <i>et al.</i> (2015)	Evaluation of a Context-Aware Voice Interface for Ambient Assisted Living : Qualitative User Study vs. Quantitative System Evaluation	Présentation des résultats d'une expérience avec des personnes âgées et des personnes malvoyantes interagissant avec une maison intelligente à l'aide de commandes vocales (projet Sweet-Home)	Personnes âgées et personnes malvoyantes	Parole	Son	Commande vocale	Habitat intelligent, domotique
[18]	Derboven <i>et al.</i> (2014)	Designing Voice Interaction for People with Physical and Speech Impairments	Découvrir comment les utilisateurs ayant des troubles physiques et de la parole interagissent avec l'interface vocale d'un environnement connecté (projet ALADIN)	Personnes ayant des troubles physiques et de la parole	Parole	Environnement en 3D (feedback du système affiché à l'écran)	Commande vocale	Habitat intelligent, domotique
[1]	Abdolrahmani <i>et al.</i> (2018)	“Sjri Talks at You” : An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind	Comprendre comment les personnes aveugles expérimentent les assistants personnels vocaux et les difficultés rencontrées lors de la tentative d'accès à ces technologies (Siri, Google Assistant, Amazon Echo, Google Home)	Personnes malvoyantes	Parole	Son, affichage sur écran	Conversationalnel	Assistant intelligent (appareils mobiles ou enceinte connectée)
[4]	Azenkot <i>et al.</i> (2013)	Exploring the Use of Speech Input by Blind People on Mobile Devices	Explorer les modèles et les défis de l'interaction vocale avec des appareils mobiles par des personnes malvoyantes (outil d'accessibilité VoiceOver sur iOS)	Personnes malvoyantes	Parole, toucher, gestes	Son, affichage sur écran	Inconnu	Outil d'accessibilité pour smartphones et tablettes
[83]	Wolters <i>et al.</i> (2010)	Making It Easier for Older People to Talk to Smart Homes : The Effect of Early Help Prompts	Étudier l'effet du placement des messages d'aide sur l'interaction des utilisateurs âgés avec l'interface vocale d'un environnement intelligent (projet INSPIRE)	Personnes âgées	Parole	Son	Conversationalnel	Habitat intelligent, domotique
[24]	Furqan <i>et al.</i> (2017)	Learnability Through Adaptive Discovery Tools in Voice User Interfaces	Fournir un résumé d'une approche de conception visant à résoudre certains des problèmes d'apprentissage dans les interfaces vocales (projet DiscoverCal)	Inconnu	Parole	Son, affichage sur écran	Inconnu	Application bureau
[48]	Myers <i>et al.</i> (2018)	Patterns for How Users Overcome Obstacles in Voice User Interfaces	Identifier les principales catégories d'obstacles dans les problèmes d'interface vocale et les types de tactiques utilisées par les utilisateurs pour les surmonter (projet DiscoverCal)	Inconnu	Parole	Son, affichage sur écran	Inconnu	Application bureau
[86]	Zeng <i>et al.</i> (2017)	End User Security and Privacy Concerns with Smart Homes	Découvrir comment les utilisateurs interagissent avec leur habitat intelligent et comprendre leurs attitudes, attentes et actions en matière de sécurité et de respect de la vie privée	Inconnu	Inconnu	Inconnu	Inconnu	Habitat intelligent, domotique

Tableau 3.2 – Études de cas relatives à l'interaction vocale et les environnements intelligents (partie 1)

Ref.	Auteur(année)	Titre	Objectif	Groupe ciblé	Modalité d'entrée	Modalité de sortie	Style d'interaction	Contexte
[77]	Vipperla <i>et al.</i> (2009)	Speech Input from Older Users in Smart Environments : Challenges and Perspectives	Examiner l'effet de l'âge sur les performances de la reconnaissance vocale en termes de facteurs acoustiques et linguistiques	Personnes âgées	Parole	Son	Inconnu	Habitat intelligent, domotique
[11]	Calleja <i>et al.</i> (2009)	Designing smart home interfaces for the elderly	Définir les bases de la conception et du développement d'un système de dialogue multimodal pour assister les personnes âgées dans leurs tâches quotidiennes	Personnes âgées	Inconnu	Inconnu	Inconnu	Habitat intelligent, domotique
[36]	Kowalski <i>et al.</i> (2019)	Older Adults and Voice Interaction : A Pilot Study with Google Home	Examiner le potentiel des assistants vocaux pour certaines catégories de personnes âgées dans le contexte de l'habitat intelligent	Personnes âgées	Parole	Son	Conversationalnel	Assistant personnel intelligent (enceinte connectée)
[85]	Zajicek <i>et al.</i> (2004)	Speech interaction for older adults	Explorer l'utilisation de la technologie vocale interactive avec un système de réservation par des personnes âgées	Personnes âgées	Parole	Son	Conversationalnel	Système de réservation par la parole
[40]	López-Cózar <i>et al.</i> (2010)	Multimodal Dialogue for Ambient Intelligence and Smart Environments	Examen de certaines questions importantes liées à la conception, la mise en œuvre, la performance et l'évaluation de systèmes de dialogue multimodaux pour l'intelligence ambiante et les environnements intelligents	Inconnu	Parole, toucher, modalités traditionnelles	Son, affichage sur écran	Inconnu	Habitat intelligent, domotique
[41]	Luger <i>et al.</i> (2016)	"Like Having a Really bad PA" : The Gulf between User Expectation and Experience of Conversational Agents	Compréhension des facteurs influençant l'expérience utilisateur des assistants conversationnels (Siri, Google Now, Cortana)	Inconnu	Parole, toucher	Son, affichage sur écran	Conversationalnel	Assistant personnel intelligent (appareils mobiles)
[55]	Porcheron <i>et al.</i> (2017)	"Do Animals Have Accents?" : Talking with Agents in Multi-Party Conversation	Examiner l'utilisation des assistants personnels intelligents dans une conversation anodine et dégager le caractère de leur utilisation (Siri, Google Now, Cortana)	Inconnu	Parole, toucher	Son, affichage sur écran	Conversationalnel	Assistant personnel intelligent (appareils mobiles)
[25]	Gemmeke <i>et al.</i> (2013)	Self-taught assistive vocal interfaces : An overview of the ALADIN project	Donner un aperçu du projet ALADIN, qui vise à développer une interface vocale d'assistance pour les personnes ayant une déficience physique	Personnes ayant des troubles physiques et de la parole	Parole	(Environnement en 3D, feedback du système affiché à l'écran)	Commande vocale	Habitat intelligent, domotique
[47]	Myers <i>et al.</i> (2019)	The Impact of User Characteristics and Preferences on Performance with an Unfamiliar Voice User Interface	Analyse des caractéristiques et des préférences des utilisateurs sur la façon dont ils interagissent avec un calendrier basé sur une interface vocale (projet DiscoverCal)	Inconnu	Parole	Son, affichage sur écran	Inconnu	Application bureau
[15]	Cowan <i>et al.</i> (2017)	What Can I Help You With ? : Infrequent Users Experiences of Intelligent Personal Assistants	Enquêter sur la façon dont les utilisateurs occasionnels interagissent avec l'assistant personnel intelligent d'Apple (Siri)	Personnes qui n'utilisent pas fréquemment les assistants personnels intelligents	Parole, toucher	Son, affichage sur écran	Conversationalnel	Assistant personnel intelligent (appareils mobiles)
[69]	Stifelman <i>et al.</i> (2013)	Designing Natural Speech Interactions for the Living Room	Présentation des résultats de la conception et de l'étude d'un prototype permettant de rechercher des films et des émissions de télévision avec une interface vocale	Inconnu	Parole	Son, affichage sur écran	Conversationalnel	Centre multimédia

Tableau 3.3 – Études de cas relatives à l'interaction vocale et les environnements intelligents (partie 2)

Sur l'ensemble des études, treize articles contiennent des expérimentations ou interviews impliquant des populations à besoins spécifiques. Parmi ces populations, les personnes âgées sont les plus représentées ($n=7$), suivies par les personnes malvoyantes ($n=4$), les personnes souffrant de troubles moteurs ($n=3$), et pour finir celles éprouvant des difficultés avec l'usage de la parole ($n=2$). La prépondérance des études basées sur une population âgée peut s'expliquer par le fait qu'il s'agisse de la population à besoins spécifiques la plus représentée à l'échelle mondiale. Vieillir étant inéluctable, chacun est potentiellement amené à faire partie de cette population. De plus, la vieillesse implique parfois plusieurs troubles partagés par d'autres populations à besoins spécifiques tels qu'une baisse de l'acuité visuelle, des difficultés motrices, ou une détérioration des capacités cognitives[60].

La plupart des articles, à l'exception de deux études[86, 11], traitent de l'interaction vocale. Parmi les modalités d'interaction retenues, la parole est la méthode de saisie principale. Cinq études utilisent le toucher conjointement à la parole, un choix de prédilection pour les personnes à mobilité réduite[53]. En sortie, une majorité des études ($n=12$) privilégient la synthèse vocale et l'affichage sur un écran. Bien que cette décision pallie les problèmes inhérents à la nature invisible de l'interaction vocale, elle ne prend pas en compte les aptitudes des populations malvoyantes et nécessite d'attirer le regard de l'utilisateur, annulant l'un des intérêts de l'interaction vocale par la même occasion. Seulement six études ont recours exclusivement à la synthèse vocale pour informer l'utilisateur.

Les contextes d'utilisation sont variés. Le contrôle d'un environnement intelligent ou connecté est la principale préoccupation parmi les différentes études ($n=9$). S'ensuivent l'utilisation d'assistants personnels intelligents dans un contexte nomade ($n=4$) ou sédentaire ($n=3$), les applications mobiles ($n=2$) et de bureau ($n=3$), la réservation par téléphone avec un répondeur automatique ($n=1$), et le contrôle d'un centre multimédia ($n=1$). L'intérêt des environnements connectés dans les études relatives à l'interaction vocale peut être dû au fait que la parole est l'une des modalités les plus utilisées pour contrôler les dispositifs dépourvus d'écrans[56].

3.2.3 Études de cas : Principales caractéristiques

À la lecture des différentes études, il apparaît que la majorité des préoccupations des utilisateurs se rapportent à des problèmes d'utilisabilité. La diversité et l'évolution des aptitudes des utilisateurs ne sont pas des facteurs prépondérants à l'émergence de ces problèmes. L'explication la plus vraisemblable est que ce ne sont pas des problèmes propres aux populations à besoins spécifiques : il existe des similitudes dans l'expression des difficultés rencontrées quelque soit le groupe ciblé. Cependant, la conception des systèmes basés sur l'interaction vocale ignore la variation des aptitudes de l'ensemble des utilisateurs et pallie les problématiques d'utilisabilité par des mécanismes non-adaptés aux populations à besoins spécifiques. La manière dont l'utilisateur interagit avec le système est particulièrement concernée par ce constat. Les principales caractéristiques suivantes regroupent la plupart des problèmes d'utilisabilité exprimés.

- **Contrôle de l'interaction avec le système :** L'interaction vocale est codifiée. Il y a un rythme, une syntaxe et un vocabulaire spécifiques à respecter. Face à un système opaque et rigide, la liberté de l'utilisateur est limitée. La sensation de ne pas avoir la main-mise sur l'interaction est l'une des principales préoccupations exprimées par les utilisateurs dans l'ensemble des articles et plus particulièrement pas les populations à besoins spécifiques.

Lorsque l'utilisateur est dans l'incapacité d'utiliser la parole pour interagir avec le système ou s'il ne parvient pas à ses fins efficacement, le recours à d'autres modalités est souhaité. L'efficacité peut être rencontrée de plusieurs manières telles que l'adaptation du système aux caractéristiques de l'utilisateur ou une meilleure flexibilité aux différents style d'interaction.

- **Nature invisible de l’interaction vocale** : L’un des avantages de l’interaction vocale est la reconnaissance omnidirectionnelle de la parole[38]. Cet avantage est aussi l’un de ses plus gros inconvénients selon les différentes études, indépendamment de l’appartenance à une population fragilisée. Le fait de s’affranchir d’un retour visuel restreint grandement les possibilités d’informer l’utilisateur sur les actions du système. Ainsi, une majorité des articles mentionnent un manque de retour d’information.

La nature invisible de l’interaction vocale complique la découverte et l’apprentissage de commandes vocales[84]. Les besoins d’aide et de mécanismes facilitant la découverte font pourtant partie des problématiques les moins citées.

- **Gestion des erreurs du système et de l’utilisateur** : L’absence d’affichage impacte l’efficacité avec laquelle l’utilisateur corrige ses erreurs et apprend de celles-ci. Plusieurs articles font état d’incompréhensions et de difficultés à surmonter les erreurs lorsqu’elles se présentent. L’incapacité du système à empêcher ces erreurs de se produire et de fournir des moyens appropriés pour les corriger est également mentionnée. Les personnes malvoyantes sont plus spécifiquement impactées par la complexité de la correction des erreurs.

Les avis recueillis relatifs à l’expérience utilisateur ne sont pas plus marqués entre les différents groupes ciblés à l’exception des craintes exprimées par les populations à besoins spécifiques. Toutefois, il ressort que l’interaction vocale joue un rôle important pour ce qui est de l’assistance dans les tâches quotidiennes pour ces populations. Certains points apparaissant comme étant plus importants alors que d’autres dépendent de la gravité des troubles affectant l’utilisateur. Quoi qu’il en soit, les considérations des populations à besoins spécifiques visent à améliorer leur expérience plutôt qu’à l’agréments d’artifices inutiles.

- **Intelligence du système** : Contrairement aux modalités d’interaction traditionnelles, l’interaction vocale nécessite un certain niveau d’intelligence pour traduire les paroles de l’utilisateur en actions perceptibles par celui-ci. Hormis la reconnaissance automatique de la parole, la connaissance du contexte et les techniques d’apprentissage automatique facilitent les prises de décision du système. Nombre d’utilisateurs regrettent que cette intelligence ne soit pas mise à profit pour mieux cerner leurs attentes et s’adapter à leurs besoins spécifiques et dynamiques.

Les populations à besoins spécifiques se montrent plus inquiètes que les autres à l’égard de la place prise par l’interaction vocale dans leur quotidien et des possibles dérives.

- **Humanité du système** : L’entretien d’un dialogue à voix haute avec une entité invisible ayant des caractéristiques humaines interpelle les utilisateurs. La métaphore humaine pousse les utilisateurs à adopter un style conversationnel avec le système et pose la question de la pertinence d’un tel style d’interaction selon l’usage.
- **Gestion de la sécurité et de la fiabilité du système** : Peu d’articles traitent des préoccupations des utilisateurs quant à leur sécurité, la collecte de leurs données privées et l’usage de l’interaction vocale dans les lieux publics. Des préoccupations qui varient selon les populations et la gravité de leur état de santé : les populations à besoins spécifiques acceptent plus facilement les compromis de sécurité pour plus de commodités. La nature opaque des interfaces vocales tend à masquer le processus de décision du système et remet en question l’intelligence de celui-ci. Quelques articles abordent la méfiance des utilisateurs dans l’interaction vocale et la fiabilité d’un tel système. Les risques que celui-ci soit inopérant et l’absence de solutions alternatives contribuent à nourrir leurs craintes.
- **Intégration du système dans son environnement** : En s’inscrivant dans un environnement connecté, le souhait des utilisateurs est de pouvoir utiliser l’interaction vocale

pour contrôler un éventail d'appareils et d'applications toujours plus large. Le désir d'intégration et d'interopérabilité est exprimé par tous les groupes mais semble plus profiter aux populations à besoins spécifiques.

Principales préoccupations	Heuristiques de Nielsen	Heuristiques d'Arhippainen
Contrôle de l'interaction avec le système	User control and freedom Flexibility and efficiency of use	Ensure usability
Gestion des erreurs	Help users recognize, diagnose, and recover from errors	
	Error prevention	
Nature invisible de l'interaction vocale	Visibility of system status	
	Recognition rather than recall	
	Help and documentation	
Intelligence du système		Respect the user
		Surpass the user's expectations
		Design the product or service to fit the intended contexts
Gestion de la sécurité et de la fiabilité du système		Respect the user's privacy and security
Intégration du système dans son environnement	Aucune	Aucune
Humanité du système	Match between system and the real world	

Tableau 3.4 – Correspondance entre les préoccupations des utilisateurs et les heuristiques existantes

3.2.4 Lien avec les heuristiques existantes

Dans leur article[63], Rusu *et al.* proposent une méthodologie afin de rédiger des heuristiques destinées à évaluer l'utilisabilité d'un système. Leur méthodologie comporte six étapes. Dans l'ordre, ces étapes sont : réaliser un travail exploratoire du domaine d'application, rédiger une description des principales caractéristiques relevées, faire le lien entre les caractéristiques et des heuristiques existantes, formaliser l'ensemble des heuristiques proposées selon un template défini, valider les nouvelles heuristiques face aux heuristiques traditionnelles à l'aide de cas d'étude, et enfin, raffiner les heuristiques selon les résultats obtenus. Si aucune heuristique n'existe pour le domaine considéré, il convient d'utiliser les dix heuristiques de Nielsen[28].

Le travail exploratoire a permis de déterminer les principales caractéristiques ou problématiques rencontrées par les utilisateurs. Ceux-ci ont montré autant d'intérêt pour les obstacles relatifs à l'utilisabilité du système qu'aux problèmes liés à l'expérience utilisateur. Les différentes préoccupations d'utilisabilité sont mises en relation avec les dix heuristiques de Nielsen tandis que les questions d'expérience utilisateur reposent sur les dix heuristiques d'Arhippainen[2] (voir tableau 3.4).

3.2.5 Définition des heuristiques

Sur base des principales caractéristiques canalisant les réactions des utilisateurs, les heuristiques existantes sont redéfinies pour correspondre au contexte de l'interaction vocale. La

définition des heuristiques suit le template proposé par Rusu *et al.*[63] : un identifiant, le nom de l'heuristique, une définition, des explications complémentaires, un exemple pratique, une liste de bénéfices liés au respect de l'heuristique et une liste de problèmes potentiels en cas de mauvaise application. Afin de garder une cohérence dans les exemples, ceux-ci sont inspirés des solutions proposées par Amazon pour son interface vocale Alexa.

1. Facilité d'utilisation

- (a) (VOCH1)**Contrôle sur le système** : Le système devrait laisser le contrôle de l'action à l'utilisateur et ne pas l'enfermer dans un modèle strict. L'utilisateur devrait bénéficier d'une grande liberté d'action. Le système devrait garantir l'essentiel de son interaction à l'aide de la parole. Dans le cas où un support visuel est utilisé, le système devrait en limiter l'usage au strict nécessaire.

Explications : À défaut de présenter un cadre contraignant l'interaction vocale, les utilisateurs sont amenés à interagir de manière variée en utilisant différentes approches. Prendre en compte cette variété et proposer différentes options de contrôle aux utilisateurs permet de faciliter les échanges avec l'interface vocale et de garder la maîtrise sur les comportements du système. Les changements de modalités en cours de dialogue sont à proscrire afin d'éviter de nuire à la nature "main-libre" de l'interaction vocale.

Exemple : Les appareils Echo Dot disposent de boutons physiques permettant d'éveiller Alexa, de couper les microphones et de régler le volume sonore.

Bénéfices :

- Facilité d'interaction
- Maîtrise de l'interaction
- Promotion d'un usage multi-tâches

Problèmes :

- Interaction rigide
- Manque de contrôle sur l'interface

- (b) (VOCH2)**Visibilité du système et retour d'information** : Le système devrait garder l'utilisateur informé de ses actions à l'aide d'un retour approprié dans un temps raisonnable. L'information devrait être présentée à l'utilisateur en prenant en compte ses aptitudes.

Explications : L'un des problèmes de l'interaction vocale est la difficulté de savoir si les actions du système correspondent bien aux intentions de l'utilisateur. L'interprétation des énoncés peut conduire à des résultats inattendus, auquel cas tenir informé l'utilisateur de l'état du système et du traitement des requêtes contribue à une meilleure compréhension de son fonctionnement. En outre, les informations peuvent ne pas être dispensées de manière accessible pour les populations à besoins spécifiques. En tenant compte de leurs aptitudes il est donc possible de sélectionner la modalité adéquate pour informer l'utilisateur.

Exemple : Au terme de l'exécution d'une requête, Alexa diffuse un message de confirmation ou un son distinctif, selon les préférences de l'utilisateur, pour le rassurer sur l'accomplissement de sa requête.

Contre-exemple : Les appareils Echo Dot arborent un cercle lumineux indiquant à l'utilisateur certains états du système (à l'écoute, en cours de configuration, problèmes de connexion ou techniques). Ces indications visuelles manquent d'accessibilité pour les personnes malvoyantes.

Bénéfices :

- Amélioration de la compréhension du système
- Gain en confiance dans l'interaction

Problèmes :

- Problèmes d'accessibilité visuelle
- Mauvais modèle mental de l'interaction vocale

- (c) (VOCH3)**Prévention des erreurs** : Le système devrait être conçu de sorte à éviter les erreurs perceptibles par l'utilisateur.

Explications : Une grande partie des frustrations des utilisateurs provient de la répétition des erreurs de compréhension du système. Trois types d'incompréhension sont à distinguer : la faiblesse de la reconnaissance automatique de la parole n'arrivant pas à détecter correctement les mots énoncés (bruits ambiants, distance avec le microphone, etc.), les requêtes mal formulées par l'utilisateur (syntaxe et/ou vocabulaire inadéquats), et les intentions de l'utilisateur dépassant les capacités du système. L'amélioration de la reconnaissance automatique de la parole, la flexibilité de la syntaxe et du vocabulaire ainsi que la compréhension des limites de l'interaction vocale par l'utilisateur permettent de réduire bon nombre des erreurs rencontrées. Il arrive cependant que le système peine à traduire les intentions des utilisateurs malgré une formulation correcte. Pour cela, les techniques d'apprentissage automatique sont d'une aide précieuse.

Exemple : Grâce à l'apprentissage actif ("Active Learning"), Alexa est capable de réduire significativement les erreurs de compréhension en annotant les mots les plus significatifs et en leur accordant un poids dans la prise de décision[10].

Bénéfices :

- Gain en efficacité

Problèmes :

- Attentes revues à la baisse
- Frustrations des utilisateurs

- (d) (VOCH4)**Reconnaissance des erreurs et mécanismes de correction** : Le système devrait informer l'utilisateur des erreurs rencontrées et lui proposer un moyen efficace d'y remédier.

Explications : En cas d'erreur, il appartient au système d'informer clairement l'utilisateur sur le problème rencontré et sur les possibilités qui lui sont offertes pour y remédier. Bien souvent la répétition est la stratégie adoptée par les utilisateurs. Néanmoins d'autres moyens existent épargnant à l'utilisateur de devoir se répéter inlassablement tels que les réparations conversationnelles ou les réparations initiées par le système via le dialogue. Dans tous les cas, cibler le problème et renseigner l'utilisateur est primordial pour une réparation efficace.

Exemple : La correction des incompréhensions du système par le dialogue permet à Alexa de clarifier les intentions de l'utilisateur à l'aide d'un tour supplémentaire. Par exemple, en demandant d'éteindre un appareil, Alexa peut demander de répéter de quel appareil il s'agit en cas d'incompréhension sans avoir à énoncer la requête complète.

Bénéfices :

- Gain en efficacité

Problèmes :

- Frustration liée à la difficulté de correction

- Correction privilégiée via une autre modalité
- (e) (VOCH5)**Stimulation de la découverte et de l'apprentissage** : Le système devrait être en mesure de supporter la découverte et de favoriser l'apprentissage de commandes vocales de la manière la plus simple et efficace qu'il soit.
- Explications : À défaut de bénéficier d'un support visuel, l'interaction vocale peut ne pas être intuitive et sembler opaque pour l'utilisateur. En proposant des solutions pour informer l'utilisateur sur les fonctionnalités existantes et les requêtes affiliées, le système s'assure que l'utilisateur dispose d'un modèle mental correct de l'interaction vocale. Dispenser ce savoir reste un challenge qui passe bien souvent par une brève introduction et quelques exemples mis dans le contexte d'un usage réel. Par la suite, l'utilisateur peut être informé périodiquement par le système des commandes existantes.
- Exemple : Périodiquement, les utilisateurs d'Alexa reçoivent un courriel d'information leur présentant de nouvelles commandes vocales ainsi qu'une sélection de commandes existantes à essayer.
- Bénéfices :
- Familiarisation avec le système
 - Création d'un modèle mental correct
 - Gain en utilisabilité
- Problèmes :
- Opacité de l'interface
 - Manque de mise en contexte
 - Tutoriel inadapté
- (f) (VOCH6)**Aide et documentation** : Le système devrait fournir une aide contextuelle et accessible à l'utilisateur.
- Explications : Un tutoriel et quelques exemples pour mettre en condition l'utilisateur ne suffisent pas à lui apprendre toutes les commandes et fonctionnalités du système. Un support continu est nécessaire pour l'aider lorsqu'il ne parvient pas à ses fins. Cette aide peut être consultable à la demande mais également dispensée de manière proactive en fonction des performances de l'utilisateur. Dans tous les cas, il convient de respecter les aptitudes de l'utilisateur pour fournir un soutien de manière appropriée.
- Exemple : Si une fonctionnalité n'a pas été configurée au préalable, Alexa indique à l'utilisateur quelle démarche il doit entreprendre avant de l'utiliser. Par exemple, en ajoutant une entrée à l'agenda, Alexa renseigne l'utilisateur sur la manière de lier un calendrier à son compte si cela n'a pas été fait.
- Bénéfices :
- Support d'aide pertinent
 - Gain en utilisabilité
- Problèmes :
- Aide rébarbative, inconsistante, inadéquate
 - Problèmes d'accessibilité
- (g) (VOCH7)**Modalités alternatives, efficacité et flexibilité** : Le système devrait proposer diverses modalités d'interaction alternatives et laisser le choix de la modalité à l'utilisateur. Le système devrait promouvoir l'efficacité en proposant à l'utilisateur de personnaliser l'interaction à son usage.

Explications : Un grande partie des inconvénients de l'interaction vocale peut être compensée par l'ajout d'une modalité supplémentaire en entrée ou en sortie. Évidemment, cette modalité ne doit pas supplanter l'interaction vocale mais venir la compléter. L'essentiel de l'interaction doit pouvoir être réalisé à l'aide de la parole et de la synthèse vocale. La recherche de l'efficacité de l'utilisateur doit être l'une des priorités qui peut être réalisée non seulement par la multimodalité mais également par une plus grande flexibilité ainsi que la possibilité de personnaliser l'interaction selon les aptitudes et préférences de l'utilisateur.

Exemple : L'application Alexa sur les appareils mobiles permet de contrôler les appareils connectés tels que les lumières sans avoir à énoncer de requête.

Bénéfices :

- Gain en efficacité
- Facilité d'interaction
- Personnalisation du système

Problèmes :

- Modalités alternatives privilégiées par rapport à l'interaction vocale
- Difficultés d'interaction avec le système

2. (VOCH8) **Conscience de l'utilisateur** : Le système devrait apprendre à connaître l'utilisateur. Pour assurer la satisfaction de l'utilisateur, le système devrait prendre en compte ses besoins, ses attentes et ses expériences passées. Le système devrait faciliter son utilisation en s'adaptant à ses actions, ses compétences et ses aptitudes.

Explications : Chaque utilisateur a des caractéristiques différentes influençant son expérience de l'interaction vocale. En proposant une interface unique selon l'approche de conception "One-Size-Fits-All", le risque est d'ignorer certaines populations aux besoins plus spécifiques. À la place, le système devrait se montrer adaptatif et proposer une interaction adéquate selon chacun. Il en résulte qu'un système plus souple nécessite de connaître l'étendue des besoins des utilisateurs et de mettre en oeuvre plus de moyens pour les combler. Outre les besoins et attentes, il convient de tirer parti des biais d'assimilation en proposant une interaction qui ne diverge pas de trop des codes de la concurrence. Il est inutile de réinventer la roue.

Exemple : Grâce aux techniques de Self-Learning, Alexa est capable de détecter les signaux contextuels implicites ou explicites des utilisateurs afin de détecter les interactions non-satisfaisantes ou les échecs de compréhension[10]. Alexa apprend comment répondre à ces problèmes automatiquement.

Bénéfices :

- Adaptation aux préférences et aptitudes de l'utilisateur

Problèmes :

- Sentiment de dépendance au système
- Biais d'assimilation
- Attentes élevées
- Perte de capacités cognitives par manque de pratique

3. (VOCH9) **Conscience du contexte d'utilisation** : Le système devrait prendre conscience du contexte dans lequel s'inscrit l'action de l'utilisateur, et de son environnement, pour s'y adapter.

Explications : Le contexte d'utilisation joue un grand rôle dans la proactivité du système et sa compréhension des requêtes. En connaissant le contexte, le système est à même

d'adopter un comportement adéquat et de dispenser des informations pertinentes. En matière de prévention des erreurs, la connaissance du contexte permet de solutionner les problèmes d'ambiguïté.

Exemple : Alexa est capable de conserver des données contextuelles au travers de plusieurs tours de conversation. Par exemple, en demandant s'il va pleuvoir ce jour ("Alexa, pleuvra-t-il aujourd'hui ?") et en suivant cette requête par, "Alexa, et demain ?", Alexa est capable de déterminer qu'il est toujours question du risque de pluie mais pour le jour suivant. De plus, Alexa utilise les données de localisation pour informer des risques d'averses sans avoir à préciser de lieu.

Bénéfices :

- Proactivité du système
- Prévention des erreurs de compréhension

Problèmes :

- Embarrasement social
- Incompréhension dans les milieux bruyants

4. (VOCH10)**Respect de l'intimité de l'utilisateur et amélioration de sa sécurité :**

Le système devrait respecter l'intimité de l'utilisateur en toutes circonstances et fournir des moyens de garantir sa sécurité.

Explications : L'intimité et la sécurité des utilisateurs sont souvent reléguées au second plan au profit des fonctionnalités du système. Ce compromis est difficilement acceptable. Il convient donc d'adopter des pratiques respectueuses de la confidentialité des données de l'utilisateur et de lui laisser la possibilité de contrôler l'enregistrement des conversations (p. ex. : coupure manuelle des microphones, gestion de l'historique des enregistrements, etc.). Dans les environnements connectés, l'interaction vocale peut jouer un rôle en matière de sécurité en informant l'utilisateur des risques potentiels. Par contre, la possibilité qu'un tiers puisse contrôler ce même environnement à la place de l'utilisateur principal pose de sérieuses questions sur les possibles failles de sécurité.

Exemple : L'une des fonctionnalités d'Alexa disponible aux États-Unis, Alexa Guard, permet de détecter un bris de vitre ou le déclenchement d'une alarme durant l'absence de l'utilisateur et de l'en informer par message[10].

Contre-exemple : Amazon reconnaît réécouter ultérieurement les conversations des utilisateurs avec Alexa pour améliorer son service[16].

Bénéfices :

- Préservation de l'intimité de l'utilisateur
- Prévention d'éventuels risques pour l'utilisateur

Problèmes :

- Inconscience des utilisateurs des risques potentiels liés au non-respect de la confidentialité de leurs données privées et aux failles de sécurité
- Compromis au détriment de la sécurité et de la confidentialité

5. (VOCH11)**Entretien d'une relation de confiance basée sur la fiabilité et la transparence :**

Le système devrait rassurer l'utilisateur sur son usage en agissant de manière transparente et en assurant son bon fonctionnement. En cas de défaillance, le système devrait être en mesure de proposer un moyen à l'utilisateur de parvenir à ses fins.

Explications : Le manque de confiance accordée à l'interaction vocale scelle les attentes des utilisateurs et restreint l'usage de l'interface vocale à des tâches basiques. Garantir à

l'utilisateur un fonctionnement correct, ou tout du moins fournir un retour d'information pertinent, permet à l'utilisateur de prendre confiance en l'interaction. Il convient également de s'assurer que le système soit résilient aux diverses pannes pouvant survenir. Un passage en mode dégradé ou la possibilité de revenir à une interaction physique permet à l'utilisateur de garder le contrôle sur l'interaction.

Exemple : Lors de certaines requêtes, comme la programmation d'une minuterie, Alexa confirme et répète la commande énoncée : "X minutes à partir de maintenant".

Bénéfices :

- Résilience aux pannes
- Confiance dans le système

Problèmes :

- Circonscription de l'interaction à des tâches basiques en cas de méfiance
- Inconsistance de l'interaction par manque de fiabilité

6. (VOCH12)**Correspondance avec le monde réel** : Le système devrait proposer une interaction familière et naturelle à l'utilisateur.

Explications : La correspondance avec le réel garantit à l'utilisateur de retrouver des repères pour l'aider dans l'interaction. Néanmoins, même si le réalisme de l'interaction peut être bénéfique, il existe une limite à ne pas dépasser sans quoi un malaise pourrait être ressenti⁴. Enfin, dans certains contextes d'utilisation, la correspondance au réel peut ne pas être la métaphore la plus appropriée.

Exemple : Le mode Whisper contribue à renforcer les caractéristiques anthropomorphiques d'Alexa. Lorsqu'un utilisateur chuchote une requête à Alexa, celle-ci répond également en chuchotant[10]. Cette fonctionnalité n'est toutefois disponible qu'aux États-Unis actuellement.

Bénéfices :

- Facilité d'interaction grâce aux similitudes avec le réel
- Renforcement des facteurs hédonique et ludique liés à l'interaction vocale

Problèmes :

- Malaise provoqué par une représentation trop réaliste
- Métaphore inutile selon certains contextes

7. (VOCH13)**Support de l'intégration** : Le système devrait s'intégrer dans un plus vaste environnement et offrir la possibilité d'interagir avec un maximum d'appareils ou d'applications tierces.

Explications : Dans le cadre des environnements connectés, l'interaction vocale promeut la centralisation, l'efficacité et l'indépendance des utilisateurs en se comportant tel un point de contrôle unique. Pour ce faire, l'interface vocale doit permettre d'interagir avec une large variété d'appareils et d'applications sans quoi son intérêt n'en serait plus que limité.

Exemple : Grâce au kit de développement de skills d'Alexa, les développeurs et concepteurs tiers ont la possibilité de permettre l'interaction avec leurs produits via Alexa.

Bénéfices :

- Contrôle centralisé de nombreuses solutions
- Gain en efficacité
- Promotion de l'indépendance des populations à besoins spécifiques

4. Voir "La Vallée de l'étrange"[7]

Problème :

- Inutilité de l'interface par manque d'intégration avec les autres services et appareils connectés

3.3 Évaluation par heuristiques : Résultats et interprétations

Résumé

À la faveur d'une analyse des vingt-deux articles, les préoccupations des utilisateurs ont été collectées et catégorisées selon les treize heuristiques définies auparavant afin d'en juger la pertinence. Une interprétation des résultats ainsi obtenus met en exergue les besoins, les attentes et les problèmes rencontrés par les utilisateurs. Des pistes sont proposées pour solutionner ces constatations.

3.3.1 (VOCH1) Contrôle sur le système

Résultats des études

Il est difficilement concevable qu'un système avec une interface vocale soit en mesure d'écouter, de transcrire en texte, d'interpréter et de prendre action à la moindre parole perceptible dans son environnement. Outre la difficulté de parvenir à concevoir une telle solution de manière satisfaisante pour l'utilisateur, certains jugent cette intrusion dans leur vie sociale discutable. Différentes alternatives existent comme le recours à un mot clé ou bien la pression sur un bouton activant la reconnaissance de la parole. Selon Derboven *et al.*[18], les utilisateurs seraient enclin à utiliser un mot-clé pour éviter l'interprétation d'une conversation comme une commande. La personnification des systèmes contribue à l'usage d'un nom propre pour s'adresser au système et l'activer. Vacher *et al.*[75] ont observé que les utilisateurs ne ressentaient pas le besoin de commencer leurs phrases par ce nom propre. Ainsi, le nom est parfois prononcé après l'énonciation d'une commande vocale. Dans l'étude d'Abdolrahmani *et al.*[1], les utilisateurs émettent des réserves quant à l'utilisation d'un mot d'éveil. Le système n'est pas toujours réceptif ou, au contraire, s'active de manière inopinée au beau milieu d'une conversation. En conséquence, ils préfèrent presser un bouton plutôt que d'articuler un nom propre. Dans le cas d'un système prenant place à l'intérieur du lieu de vie, l'activation inopinée est plus fréquente du fait de nombreux bruits ambiants. Le microphone de l'appareil est donc désactivé manuellement lors de rassemblements sociaux.

Vacher *et al.*[75] ont remarqué que les personnes âgées ont tendance à s'adresser au système à la manière d'une discussion naturelle entre deux individus. De nombreux mots de courtoisie sont ajoutés tels que "merci" ou "s'il-te-plaît". Les requêtes sont formulées de manière interrogative plutôt qu'injonctive. Ces découvertes sont corroborées par l'étude de Vipperla *et al.*[77]. Les utilisateurs âgés essaient d'avoir une interaction sociale avec une entité. Ils utilisent un vocabulaire plus varié et tentent de prendre l'initiative du dialogue en faisant des suggestions même avec un système dirigiste. Le souhait d'avoir l'initiative du dialogue est d'ailleurs partagé par d'autres populations selon Myers *et al.*[48]. Par contre, les observations de Gemmeke *et al.*[25] ne permettent pas de mettre en lumière une tendance. Ainsi, les styles de dialogue varient entre injonctif et anthropomorphique. Il est à noter que les injonctions sont plus techniques et requièrent un vocabulaire spécifique. La cible des injonctions varie également. Certains utilisateurs s'adressent directement aux appareils connectés ("porte, ouvre") alors que d'autres considèrent le système comme un tout agissant sur l'environnement ("[SYSTÈME], ouvre la porte").

Certaines faiblesses inhérentes à l'interaction vocale nuisent au contrôle du système. Le temps alloué à la formulation d'une commande est limité. Selon Abdolrahmani *et al.*[1], les utilisateurs trouvent cette contrainte frustrante surtout pour des commandes complexes telles que la composition d'un courriel ou l'ajout d'une entrée dans l'agenda. Autre contrainte : le non-respect de la nature "main-libre" de l'interaction vocale, un problème identifié par les utilisateurs de l'étude de Cowan *et al.*[15]. Le système invite régulièrement l'utilisateur à regarder le support visuel et à utiliser une autre modalité d'interaction. Enfin, la peur de perdre le contrôle sur son habitat et l'illusion de contrôle sur des appareils intelligents et autonomes sont des craintes exprimées par les utilisateurs des études de Vacher *et al.*[75] et de Kowalski *et al.*[36]. Pour ces derniers, il est difficile de prévoir si l'interprétation d'un ordre donné au système correspond au souhait de l'utilisateur.

Interprétation des résultats

Usage d'un mot d'éveil : Afin d'éviter l'activation inopinée du système et son intrusion intempestive dans le quotidien de l'utilisateur, l'usage d'un mot-clé ou d'un nom pour éveiller le système s'avère être une solution idéale afin de distinguer les commandes vocales des conversations naturelles[18, 57, 25]. Néanmoins, le choix de ce mot n'est pas anodin. Un mot ou un prénom trop commun risque d'activer le système accidentellement lors d'activités sociales ou de divertissement. À l'inverse, un mot trop long ou trop compliqué est difficile à mémoriser et à articuler intelligiblement.

Gestion du dialogue : Plusieurs participants se mettent dans une situation de dialogue avec le système même s'il est dépourvu de fonctionnalités conversationnelles[75]. Le système peut être perçu comme une présence dont le manque de retour frustre les utilisateurs désireux d'entretenir un échange bidirectionnel. Une gestion du dialogue, même minime, est donc nécessaire.

Initiative du dialogue : Trois stratégies d'interaction sont généralement distinguées selon qui ait l'initiative du dialogue (le système ou l'utilisateur)[40]. Si l'utilisateur a l'initiative, le système ne fait que répondre aux commandes énoncées verbalement. Une telle stratégie peut l'induire en erreur en donnant une impression de liberté. L'utilisateur pense pouvoir dire ce qu'il veut ce qui tend à produire des erreurs de reconnaissance et de compréhension du système. A contrario, l'initiative accordée au système limite les possibilités de l'utilisateur. Il ne fait que répondre aux requêtes du système dans l'ordre que celui-ci décide. Ce manque de flexibilité a au moins l'avantage de restreindre les différentes possibilités de saisie vocale de l'utilisateur. De ce fait, cette stratégie s'avère très efficace mais au détriment de l'expérience de l'utilisateur. Enfin, un mélange des deux stratégies permet à la fois à l'utilisateur et au système de prendre l'initiative dans le dialogue. Le système peut dès lors guider l'utilisateur dans ses activités. De son côté, l'utilisateur peut prendre l'initiative et être trop informatif. Le choix de l'initiative doit donc se faire avec prudence en fonction de l'usage prévu de l'interaction vocale. Ce choix peut également se faire dynamiquement et être inféré depuis le contexte d'utilisation.

Limite de temps et de complexité : Pour les personnes malvoyantes, l'interaction par la parole est une alternative plus simple que les méthodes d'interaction traditionnelles[1]. Néanmoins, la saisie vocale de commandes longues ou complexes n'est pas toujours prise en charge par l'interface vocale. Le système doit être en mesure de traiter de telles commandes pour les populations n'ayant pas d'autres alternatives satisfaisantes.

Respect de la nature "main-libre" : Le maintien de la parole comme principale modalité tant en entrée qu'en sortie doit être une priorité dans la conception d'un système intégrant une interface vocale afin que les tâches ne soient pas interrompues par un changement de modalité d'interaction[15]. L'introduction d'un élément de connaissance du contexte doit permettre au système de déterminer si un changement de modalité est nécessaire de sorte que l'interaction soit plus adaptée à la situation de l'utilisateur.

Perte de contrôle : Certaines populations, notamment les personnes âgées, ne veulent pas d'un système les conduisant vers une situation de dépendance[57]. Les utilisateurs veulent garder le contrôle sur leur environnement. Une manière d'y parvenir est de conserver l'interaction physique avec les appareils connectés conjointement à l'interaction vocale[86]. Cette solution a également les avantages d'améliorer l'interaction dans les environnements comprenant plusieurs utilisateurs et de pallier les problèmes techniques comme des pannes électriques ou de réseau.

3.3.2 (VOCH2) Visibilité du système et retour d'information

Résultats des études

Contrairement à une interface graphique, l'interface vocale est limitée en matière de visibilité et de retour d'information. Vacher *et al.*[75] ont rapporté que les participants de leur étude déplorent un manque d'information de la part du système autre que l'action à réaliser. Il n'y a aucune indication confirmant la compréhension de la commande ou son accomplissement. Un avis que partagent les utilisateurs de l'étude de Kowalski *et al.*[36]. D'après Luger *et al.*[41], la moitié de leurs participants ignorent ce dont le système est capable. Le manque de retour d'information et de visibilité des limites et capacités de l'interaction vocale sont des facteurs limitant la capacité des utilisateurs à parvenir à leurs fins. Par ailleurs, Luger *et al.* ont découvert que les utilisateurs ne se risquent pas à formuler des requêtes complexes ou sensibles en l'absence de confirmation du système. Le désir d'avoir plus d'informations est également formulée par les utilisateurs de l'étude de Myers *et al.*[47] où il apparaît que ceux désirant le plus de retour du système sont généralement les mêmes ayant mis le plus de temps à accomplir les tâches de l'expérimentation. Dans une autre étude, Myers *et al.*[48] ont observé que les participants ignorent parfois les informations fournies par le système ou les interprètent mal ce qui ne fait qu'augmenter le nombre d'erreurs.

Lorsqu'un retour d'information sous forme de synthèse vocale est prévu, les utilisateurs partagent des expériences où ce retour est trop verbeux, inutile, incomplet ou manquant de pertinence selon Abdolrahmani *et al.*[1]. Différentes manières d'avertir l'utilisateur existent. Ainsi, Portet *et al.*[57] ont rapporté que les utilisateurs souhaitent que le système émette des sons courts avant de s'exprimer afin d'y être plus attentif. Ces sons devraient varier en fonction de la nature de l'alerte. Stifelman *et al.*[69] ont réalisé une expérience avec un centre multimédia contrôlable avec la parole. Ils ont tiré profit de l'affichage pour proposer un retour d'information continu à l'écran. Les participants de leur étude apprécient d'avoir cette fonctionnalité : cela leur permet de déceler les erreurs plus rapidement et de regarder simultanément leur contenu et les informations du système. Dans la suite de leur expérimentation, le retour d'information affiché à l'écran était complété par une synthèse vocale. Le premier présentait une information contextuelle alors que le second traitait la requête en cours. Les utilisateurs estiment que la synthèse vocale fournit une assurance que le système les comprend. Ils trouvent que l'empreinte de cette solution est minime : ils n'éprouvent pas de gêne, le son n'est pas trop fort et l'information n'est pas répétitive.

Quel que soit la solution adoptée, celle-ci se doit d'être accessible. Un utilisateur de l'étude d'Abdolrahmani *et al.*[1] met en évidence le fait que les repères visuels utilisés par les appareils connectés n'ont pas d'alternative pour les populations malvoyantes. Pradhan *et al.*[58] rapportent que deux utilisateurs ont fait mention du même problème dans leur étude.

Interprétation des résultats

Manque de retour du système : Il est essentiel que le système informe l'utilisateur à la fois de son statut (p. ex., à l'écoute, en attente d'une commande, en panne, micro désactivé, etc.) mais également de toutes les informations propres à l'exécution d'une commande hormis le

résultat de la commande en question (p. ex., avertissement de l’accomplissement de la tâche, confirmation de la compréhension de la commande, mécanismes de gestion des erreurs, etc.). Un retour d’information pertinent doit permettre à l’utilisateur d’améliorer son modèle mental de l’interaction vocale et l’apprentissage des commandes.

Solutions de retour : Le retour d’information du système n’est pas toujours approprié compte tenu du contexte d’utilisation et des préférences de l’utilisateur[1]. La personnalisation du retour doit permettre à l’utilisateur de choisir le degré de détails et la modalité désirée pour les informations transmises par le système. Plusieurs solutions existent comme des signaux sonores significatifs en plus de la synthèse vocale[40] ou bien le recours à l’affichage sur un écran[69]. Compte tenu de la nature de l’interaction vocale visant à s’affranchir d’un support visuel, il est préférable de privilégier un retour audible plutôt que visuel. Malgré cela, si la solution visuelle est retenue, il est nécessaire de prendre en compte l’espace restreint de l’affichage pour ne pas gêner le contenu[69, 14].

Accessibilité du retour d’information : Le retour d’information devrait se faire en fonction des aptitudes de l’utilisateur. Les solutions visuelles, telles que l’affichage sur un écran ou les signaux lumineux sur les appareils connectés, manquent d’alternatives accessibles aux personnes malvoyantes[1]. Un retour audible est donc à privilégier pour ces populations. De brèves impulsions sonores (earcons) pourraient être utilisées pour les notifications et les alertes dispensées par le système. De même, s’il apparaît évident de ne pas informer l’utilisateur après l’accomplissement de chaque action, les personnes malvoyantes et les autres populations[69] y trouvent l’assurance de la bonne compréhension et exécution de leurs commandes. Un facteur important, pour les personnes âgées, est de limiter la taille des messages et instructions synthétisés par le système[85]. Les messages peuvent être courtois mais contiennent au plus un élément d’information ou demandent une seule donnée à la fois à l’utilisateur.

3.3.3 (VOCH3) Prévention des erreurs

Résultats des études

Selon Myers *et al.*[48], lorsque le système n’est pas en mesure de répondre favorablement à une commande vocale, la source d’erreur peut être aussi bien l’utilisateur que le système en lui-même. Dans le cas de l’utilisateur, il peut s’agir d’une commande mal structurée ou bien que le système ne supporte pas l’action désirée. Dans l’autre cas, le système a soit mal interprété la commande prononcée, soit il ne l’a pas comprise. Ce sont des erreurs liées au traitement automatique du langage naturel. Porcheron *et al.*[55] ont observé des difficultés du système à traiter les homonymes et synonymes. Pour Myers *et al.*[48], ce sont les consonnes ’N’ et ’M’ qui sont confondues. Dans l’étude d’Abdolrahmani *et al.*[1], les participants font remarquer les faiblesses du système dans des milieux bruyants. Vacher *et al.*[75] ont remarqué une correspondance entre l’âge des utilisateurs et le taux d’erreurs de traitement. Plus les personnes sont âgées et plus le taux est élevée. Une découverte confirmée par Vipperla *et al.*[77] qui mettent en cause des facteurs à la fois linguistiques et acoustiques. Malgré cela, les utilisateurs sont confiants quant à l’évolution de la technologie pour corriger les erreurs d’interprétation et de compréhension de l’interaction vocale comme le font remarquer Myers *et al.*[48].

Les connaissances techniques des utilisateurs ont une influence sur leur comportement et leurs attentes vis-à-vis de la technologie. Selon les observations de Luger *et al.*[41], les utilisateurs ayant peu de connaissances techniques ont des attentes initiales dépassant les capacités du système. Lorsqu’une tâche échoue, ils jugent le système statique et incapable d’apprendre. De ce fait, ils se montrent moins expérimentaux et s’attribuent les erreurs du système. A contrario, les utilisateurs avec un bagage technique ont des attentes modestes et sont plus enclins à accepter les erreurs du système. Dans ce cas, ils sont plus à même de déterminer la source du

problème et d'utiliser différents moyens pour arriver à leurs fins. Cowan *et al.*[15] ont identifié plusieurs techniques différentes utilisées par les participants pour pallier les erreurs tels que l'hyperarticulation, une élocution plus lente ou une formulation spécifique. L'hyperarticulation est également citée par Myers *et al.*[48] comme étant la tactique de prédilection pour minimiser les erreurs du système suivie par l'ajout d'informations complémentaires, la simplification, et la reformulation.

Certains utilisateurs peuvent se montrer intransigeants en cas d'erreurs répétées. Callejas *et al.*[11] ont rapporté que cinq pour cent des interviewés refusent d'utiliser à nouveau un système réitérant les mêmes erreurs. Les observations de Luger *et al.*[41] corroborent l'intransigeance des utilisateurs. Si le système se trompe souvent (entre deux et six fois), les attentes des utilisateurs sont fixées et ils se restreignent à des tâches basiques. Tout comme les premiers succès renforcent l'usage, les premiers échecs limitent l'utilisation.

Interprétation des résultats

Comportements des utilisateurs pour limiter les erreurs : Globalement, les utilisateurs acceptent les erreurs du système. Cette tolérance a tout de même ses limites qui, une fois franchies, conditionnent les interactions futures. Les utilisateurs ont tendance à favoriser l'hyperarticulation afin de diminuer les risques d'incompréhension ou de mauvaise traduction de leurs intentions. L'hyperarticulation est également utilisée ainsi que la simplification, la reformulation et l'ajout d'informations supplémentaires lorsque l'utilisateur énonce verbalement une commande mal structurée ou non-supportée par le système[48]. Lorsqu'un support d'aide visuel est prévu, les utilisateurs n'y ont pas toujours recours[48, 14], pas plus qu'ils ne font l'effort de se souvenir des commandes correctement formulées précédemment. Pour éviter que les erreurs ne deviennent un frein à l'adoption de l'interaction vocale, il est judicieux que le système veille à ne pas frustrer l'utilisateur et à proposer des stratégies limitant la répétition des erreurs autres qu'une aide consultable sur un autre support.

Pallier les faiblesses du traitement automatique du langage : Même en cas de compréhension partielle, le système est enclin à prendre action au risque de mal traduire les intentions de l'utilisateur. Plusieurs manières existent pour prendre en charge les commandes ambiguës ou incomplètes[18]. Premièrement, le système peut tirer profit d'une intelligence contextuelle pour inférer de l'usage courant et de la situation de l'utilisateur l'action appropriée à entreprendre. Deuxièmement, le design d'interaction peut fournir des moyens supplémentaires de résoudre les commandes ambiguës et incomplètes en ayant recours à une tierce interface accessible via d'autres modalités. Troisièmement, la mise en place d'un style d'interaction conversationnel peut permettre au système d'exiger des informations complémentaires à l'utilisateur avant d'exécuter la commande. Par exemple, les stratégies de confirmations permettent de régler les problèmes d'incertitudes[40]. La stratégie de confirmation explicite implique que le système génère un tour de dialogue supplémentaire pour confirmer la donnée obtenue du tour précédent. L'inconvénient de cette méthode de confirmation est que le dialogue a tendance à être long en raison des tours de confirmation supplémentaires, ce qui rend l'interaction moins efficace et répétitive. La stratégie de confirmation implicite ne requiert pas de tour supplémentaire car le système inclut les données à confirmer dans le prochain tour de dialogue pour obtenir d'autres informations de l'utilisateur. Dans ce cas, il est de la responsabilité de l'utilisateur de procéder à une correction s'il constate une donnée erronée. quatrièmement, une écoute continue du flux de parole peut aider le système à collecter des informations contextuelles pertinentes avant que l'utilisateur n'interagisse avec[55]. De la sorte, le système peut utiliser les informations contextuelles pour en déduire la bonne action à entreprendre. Le désavantage de cette dernière solution est qu'elle sacrifie l'intimité de l'utilisateur pour améliorer la compréhension de ses intentions. Une dernière solution consiste à s'inspirer des fonctionnalités d'auto-complétion

pour prendre en charge la formulation de requêtes sans que l'utilisateur n'ait besoin de se rappeler avec exactitude de la syntaxe nécessaire à une commande réussie[55].

3.3.4 (VOCH4) Reconnaissance des erreurs et mécanismes de correction

Résultats des études

Il n'est pas toujours possible de prévenir les erreurs. Informer l'utilisateur qu'une erreur s'est produite et lui fournir les moyens de la corriger renforce l'utilisabilité du système. D'après Portet *et al.*[57], les utilisateurs préfèrent que le système leur demande de répéter la commande vocale plutôt que de leur proposer un éventail de possibilités parmi lesquelles choisir. La répétition est la technique de prédilection des utilisateurs. Elle est citée dans l'interview de Callejas *et al.*[11] et a été observée par Porcheron *et al.*[55]. Dans l'étude de ces derniers, les participants ont rarement renoncé suite à une erreur système. S'ils n'ont pas recours à la répétition, les utilisateurs raffinent leur commande en ajoutant des informations ou en les corrigeant. À la manière d'une conversation entre êtres humains, ils tentent une réparation conversationnelle impliquant une connaissance du contexte du système, par exemple, "heu non, je voulais dire mardi soir !" après s'être trompé de jour du rendez-vous. Les participants de l'étude de Derboven *et al.*[18] ont exhibé le même comportement.

La reconnaissance des erreurs et leur correction sont jugées difficiles par les utilisateurs de l'étude d'Abdolrahmani *et al.*[1]. Une frustration partagée par les personnes malvoyantes comme l'ont constaté Azenkot *et al.*[4]. Dans le contexte de la rédaction d'un courrier à l'aide d'une interface vocale, il apparaît que quatre-vingt pour cent du temps passé à la rédaction est consacré à la reconnaissance et correction des erreurs. Lorsque ces personnes malvoyantes rédigent le même courriel au clavier, ils n'y passent plus que neuf pour cent de leur temps. Dès lors, ils préfèrent dicter leur texte oralement et le corriger via une interface physique.

Interprétation des résultats

Correction des erreurs : La difficulté d'identification et de correction des erreurs a un impact négatif sur les performances de l'utilisateur. Dès lors, l'interaction vocale est mieux adaptée aux messages courts et occasionnels[4]. La facilitation du processus de correction du contenu et de la syntaxe des énoncés devrait permettre de rendre l'usage de la parole plus efficace et plus polyvalente.

Techniques de correction : Outre l'invitation à répéter l'ordre énoncé, d'autres mécanismes de correction sont envisageables. La majorité de ces mécanismes reposent sur la pro-activité du système[55]. Le système peut informer l'utilisateur des raisons pour lesquelles sa requête a échoué ou suggérer comment la peaufiner. Il peut être également attentif aux réparations conversationnelles afin de corriger à la volée, ou avec la confirmation de l'utilisateur, la commande vocale dictée auparavant. L'utilisateur évite donc de devoir répéter l'entièreté de sa requête. Pour finir, le système peut demander à l'utilisateur de reformuler sa requête avec d'autres mots ou d'épeler un mot particulier non-reconnu.

La conception continue propose que le système répète mot pour mot, ou affiche sur un écran, la transcription en texte des commandes prononcées[69, 48]. Cette méthode a un effet bénéfique sur le modèle mental des utilisateurs. Avec un modèle mental correct des erreurs, les utilisateurs sont capables de détecter les problèmes de traduction de leurs intentions et d'en cerner les causes. En conséquence, ils peuvent adopter la tactique la plus adéquate pour corriger l'erreur parmi celles vues dans la prévention des erreurs.

3.3.5 (VOCH5) Stimulation de la découverte et de l'apprentissage

Résultats des études

La découverte et l'apprentissage des commandes vocales n'est pas la préoccupation principale des utilisateurs des différentes études. Il y a deux raisons probables à ce constat : soit les participants sont informés de la syntaxe, du vocabulaire et des possibilités du système au préalable, soit la plupart de ces participants ont un modèle mental de l'interaction vocale semblable à celui du système. Cependant quelques études font mention de difficultés quant à l'apprentissage des fonctionnalités existantes et des commandes vocales associées. Ainsi, Pradhan *et al.*[58] ont remarqué que les utilisateurs peinent à découvrir ces fonctionnalités. Ils ont exprimé le désir d'une fonctionnalité mettant en avant les différentes possibilités du système, une fonctionnalité qui pourtant existe, ce qui tend à démontrer le problème.

Dans certains cas, les utilisateurs exploitent les limites du système pour déterminer quels sont les énoncés et les intentions supportés. Myers *et al.*[47] et Corbett *et al.*[14] ont noté que les utilisateurs apprenaient l'usage d'une interface vocale par tâtonnements en multipliant les essais et les erreurs. L'aspect ludique a également une importance dans les premiers contacts. Luger *et al.*[41] ont constaté que tous les participants, à l'exception d'un seul, commencent par provoquer des réactions insolites voire humoristiques du système à l'aide de phrases bien spécifiques. Ils jugent cette expérience ludique comme une manière de se familiariser avec l'interface tout en améliorant leur compréhension des capacités du système.

L'étude de Corbett *et al.*[14] met en avant la visibilité et l'apprentissage des commandes vocales d'une application mobile. Pour ce faire, ils ont introduit plusieurs mécanismes à leur interface graphique afin d'aider les utilisateurs dans leurs tâches. Une courte introduction par la pratique a permis de mettre les participants directement dans le contexte d'utilisation de l'application et de répéter ces mêmes actions par la suite en toute indépendance. Par contre, les messages distillés tout le long de l'expérience incitant l'utilisateur à la découverte se sont révélés peu concluants. Les participants ont formulé des critiques concernant l'espace occupé par les messages affichés. Ces messages sont superposés aux éléments de l'interface compliquant la lecture de l'affichage.

Interprétation des résultats

Visibilité et apprentissage : Le faible nombre d'études relevant des problèmes d'apprentissage des commandes vocales peut s'expliquer, entre autres, par un effet d'assimilation lié aux expériences des utilisateurs avec d'autres interfaces vocales[47]. La conception d'un nouveau système basé sur l'interaction par la parole peut capitaliser sur les expériences des utilisateurs avec les principales solutions existantes pour proposer une uniformisation des requêtes et des comportements du système. Les effets d'assimilation ne sont pas toujours bénéfiques lorsque les fonctionnalités diffèrent entre les systèmes. Pour cette raison, mais également pour soutenir les utilisateurs sans expérience, des techniques de découverte et d'apprentissage sont utiles. Il s'avère qu'une brève vue d'ensemble suivie d'un apprentissage par la pratique dans le contexte d'usage du système est une initiation efficace à son utilisation future[14]. L'essentiel est de raccourcir l'introduction en passant directement à la pratique avec un exemple reflétant l'usage du système. L'efficacité des solutions de découverte est par contre mitigée. Les informations répétitives présentées dans les outils de découverte peuvent devenir redondantes et même frustrer les utilisateurs[24]. Pour les utilisateurs expérimentés, cette présentation d'informations supplémentaires n'a pas d'intérêt. De plus, conjointement à un support visuel, elle a le désavantage d'encombrer l'affichage du contenu à l'écran[14]. Une approche adaptative peut être utilisée pour personnaliser les outils de découverte en fonction des besoins des utilisateurs.

La nature invisible des interfaces vocales nécessite de trouver des moyens de visualiser ce

que le système peut accomplir et les commandes correspondantes. Néanmoins, en règle générale, l'interaction vocale est associée à des solutions graphiques[58]. L'amélioration de l'apprentissage et de la découverte pour une interaction purement non-visuelle offre des avantages non seulement aux utilisateurs malvoyants, mais aussi aux autres dans un contexte multi-tâche.

Méthodes de découverte : L'expérimentation est une pratique récurrente pour découvrir les commandes vocales. Cette méthode empirique perfectionne la compréhension du système par les utilisateurs[47]. Le système peut tirer profit des essais par tâtonnements pour mieux informer les utilisateurs des raisons pour lesquelles leur requête vocale n'a pas eu l'effet attendu. L'exploration est d'ailleurs privilégiée par rapport à un support visuel[48]. En outre, le jeu est un mécanisme essentiel pour amener les utilisateurs à des pratiques exploratoires. Néanmoins, ces pratiques tendent à attribuer des qualités anthropomorphiques au système et à renforcer les attentes des utilisateurs quant aux capacités de celui-ci[41]. L'engagement par l'humour doit donc être utilisé avec prudence et ne pas induire l'utilisateur en erreur sur l'intelligence du système.

3.3.6 (VOCH6) Aide et documentation

Résultats des études

Tout comme c'était le cas pour la stimulation de la découverte et de l'apprentissage, peu d'études se concentrent sur les fonctionnalités fournissant une aide relative à l'usage de l'interaction vocale. Les systèmes proposés par Myers *et al.*[47] et Corbett *et al.*[14] disposent d'une aide sous forme d'un menu consultable par l'utilisateur. Dans la première étude, le menu a eu un impact négatif sur l'efficacité des utilisateurs. Il est bon de noter que cette aide n'est pas contextuelle. Par contre, le système de la seconde étude contient une aide contextuelle appréciée par les utilisateurs. Leur seul regret est que les commandes vocales ne sont pas exécutables directement depuis le menu.

Contrairement aux mécanismes précédents, Wolters *et al.*[83] se cantonnent à la reconnaissance automatique de la parole et la synthèse vocale. En comparant les aides a priori et a posteriori, il s'avère que les utilisateurs âgés commettent moins d'erreurs et partagent plus de vocabulaire avec le système lorsque celui-ci oriente au préalable leurs réponses. Cet effet n'a pas pu être reproduit sur les participants plus jeunes qui ont de meilleures facultés d'adaptation.

Interprétation des résultats

Support visuel d'aide et synthèse vocale : Idéalement, l'aide fournie par le système ne devrait pas être cantonnée aux seuls supports visuels. Lorsque c'est le cas, il est nécessaire que l'aide soit relative au contexte et contruite autour des actions possibles tout en étant exécutable depuis les outils d'aide[14]. Dans le cas contraire, son efficacité tend à avoir un impact négatif sur l'efficacité des utilisateurs. Néanmoins, une telle pratique ignore les aptitudes de certaines populations, notamment les populations malvoyantes. Il faut donc un système plus accessible et privilégier la synthèse vocale comme modalité principale pour les outils d'aide. Parmi les populations les plus désireuses de profiter d'une aide du système, il y a les personnes âgées. Pour celles-ci, les messages d'aides se doivent d'être courts et intégrés au dialogue[85]. Une aide préalable à l'interaction affecte positivement leurs performances mais est sans effet pour les personnes moins âgées[83]. Les personnes âgées sont capables de se souvenir de ce qu'elles ont appris si des modèles de requêtes correctement formulées leur sont proposés tôt dans l'interaction.

3.3.7 (VOCH7) Modalités alternatives et efficacité

Résultats des études

L'interaction vocale présente de nombreux challenge mais elle a aussi des avantages indéniables prônant l'efficacité. Les participants de l'étude d'Abdolrahmani *et al.*[1] font remarquer que l'usage d'une interface vocale permet un gain de temps tout en exécutant d'autres tâches simultanément. La possibilité d'exécuter une autre tâche tout en interagissant vocalement avec le système est un avantage également cité dans les études de Pradhan *et al.*[58] et de Luger *et al.*[41]. Dans le cas des assistants personnels vocaux au sein des lieux de vie, les participants apprécient la centralisation de nombreux appareils disparates en un seul point de contrôle comme le constatent Pradhan *et al.*[58]. Les utilisateurs jugent que cela améliore leur efficacité et leur indépendance pour toute une variété de tâches. Dans des cas plus précis, comme la dictée de texte, Azenkot *et al.*[4] ont observé que les populations malvoyantes étaient plus performantes qu'en utilisant une interface traditionnelle comme le clavier. Cette performance est malgré tout contre-balançée par le taux d'erreurs et la complexité de correction.

Pour combler les lacunes de l'interaction vocale, l'ajout d'autres modalités est souvent souhaité. Même si l'usage de la parole est la modalité de préférence en entrée, nombreux utilisateurs de l'étude de Pradhan *et al.*[58] émettent le désir d'avoir d'autres moyens d'interaction comme une télécommande, un smartphone, une montre connectée, la reconnaissance gestuelle ou le contact direct avec l'appareil concerné. Ce constat est repris par les participants de l'interview de Callejas *et al.*[11]. Pour ceux-ci, un système est d'autant plus facile à utiliser s'il propose une interface multimodale.

Améliorer l'efficacité de l'interaction vocale passe également par une personnalisation du système. Dans l'étude de Cowan *et al.*[15] les participants identifient le manque de personnalisation et d'intégration comme des barrières principales à l'adoption d'un système particulier. Les utilisateurs désirent personnaliser certains éléments en se basant sur leurs besoins et la nature de la tâche à accomplir. D'après Abdolrahmani *et al.*[1], les utilisateurs veulent avoir la possibilité de personnaliser la synthèse vocale du système (p. ex., le débit d'élocution, la clarté, et l'intensité). Derboven *et al.*[18] ont conçu un système de reconnaissance automatique de la parole adaptatif. Pour les participants, l'inconvénient principal d'un tel système est qu'il faut l'entraîner au préalable pour qu'il reconnaisse de manière optimale les différentes formulations ainsi que les synonymes de l'utilisateur. Avant cet entraînement, qui peut s'avérer fastidieux, la précision du système est en deçà des attentes des utilisateurs.

Le style d'interaction adopté influe grandement sur les performances des utilisateurs. Kowalski *et al.*[36] distinguent l'interaction conversationnelle comme une solution idéale pour les personnes âgées. Cela leur permet de procéder à leur propre rythme, avec la facilité du langage naturel tout en s'affranchissant de la dépendance à un écran ou au contact avec un appareil. Les observations de Gemmeke *et al.*[25] mettent en avant le style injonctif comme étant plus efficace et plus fiable pour les populations touchées par un désordre physique. Derboven *et al.*[18] ont remarqué que lorsque les utilisateurs s'adressent au système par le nom qui lui est attribué, ils ont tendance à formuler des commandes courtes et hachurées où chaque mot est prononcé distinctement sans la moindre liaison. Cette même technique est utilisée par les participants de l'étude de Portet *et al.*[57]. La conviction des utilisateurs est qu'en utilisant principalement des mots clés, l'interaction est plus rapide et plus efficace. En matière d'initiative du dialogue, Myers *et al.*[47] ont noté que les utilisateurs préfèrent avoir l'initiative pour une question de rapidité. En outre, cela réduit la frustration de ne pas avoir le contrôle sur le système. Wolters *et al.*[83] ont découvert qu'en restreignant les personnes âgées par un choix entre plusieurs options de dialogue, celles-ci gagnent en efficacité.

Capitaliser sur le modèle mental des utilisateurs est une bonne manière d'améliorer leur efficacité. Les participants de Myers *et al.*[47] ayant une expérience avec les interfaces vocales

ont mis moins de temps à réaliser les tâches de l'expérience. À l'inverse, les connaissances en programmation n'ont pas eu d'impact significatif.

Interprétation des résultats

Efficacité de l'interaction vocale : L'usage d'une interface vocale est décrite comme un «gain de temps», réduisant le temps de travail par rapport à la saisie de données via des écrans tactiles ou à l'aide d'un clavier pour les populations spécifiques[1]. C'est aussi un point d'accès centralisé à des technologies qui leur seraient autrement inaccessibles. Pour les autres populations, l'interaction vocale permet surtout d'exécuter plusieurs tâches en même temps[41]. Même si la parole est une méthode de saisie efficace pour les personnes malvoyantes, et notamment pour la dictée de texte, le temps passé à la correction reste un frein pour une minorité d'utilisateurs. Fournir des moyens de correction de texte appropriés à ces utilisateurs est primordial pour s'assurer de leur adoption de la technologie vocale.

Multimodalité : Les limitations inhérentes à l'invisibilité d'une interface vocale comme le manque de visibilité, de retour d'information et la difficulté d'apprendre les commandes vocales, peuvent être résolus à l'aide d'autres méthodes d'interaction. Idéalement, la multimodalité ne devrait concerner que la réponse du système[40] afin d'éviter que l'ajout de plusieurs méthodes de saisie de données ne nuisent à l'utilisation de la parole comme modalité principale. Les utilisateurs pourraient être tentés de privilégier la modalité avec laquelle ils sont le plus à l'aise pour interagir avec le système au détriment de l'usage de l'interface vocale[24]. Le choix d'autres moyens de communication doit se faire de manière complémentaire à la synthèse vocale sans toutefois la remplacer. Les principales fonctionnalités doivent rester accessibles via l'interaction vocale uniquement.

Adaptation à l'utilisateur et personnalisation : L'amélioration de l'efficacité de l'utilisateur passe également par une adaptation à sa manière d'interagir avec le système. Le système peut adopter des styles d'interaction différents selon les caractéristiques de l'utilisateur[83, 47] et le contexte d'utilisation. Des erreurs multiples favorisent l'adoption d'un style conversationnel pour en limiter l'apparition ou les corriger le cas échéant. Un style injonctif est plus rapide et dans certaines situations préférable pour ne pas détourner l'attention de l'utilisateur de sa tâche principale. Les stratégies d'initiative du dialogue[40], comme vues précédemment, doivent aussi être choisies en fonction de leurs avantages et l'acceptation de leurs inconvénients. En ce qui concerne l'adaptation à la grammaire, deux profils ressortent[75]. La plupart des participants âgés dévient de la grammaire, tandis que la plupart des personnes malvoyantes la respectent. Il est probable que cela soit dû au fait que les populations malvoyantes sont familiarisées avec la reconnaissance automatique du langage. Dans tous les cas, la plupart des participants souhaitent personnaliser la grammaire malgré qu'ils n'aient rencontré aucune difficulté à utiliser le système. Ainsi, la grammaire mais également le style d'interaction doivent clairement être adaptés aux préférences de l'utilisateur. Cette adaptation peut se faire dynamiquement ou à la demande de l'utilisateur pour fournir un système plus personnalisable. A noter que pour les systèmes de reconnaissance vocale adaptatifs, les commandes qui utilisent une grammaire prévisible et un vocabulaire répétitif peuvent être mieux interprétées[18].

3.3.8 (VOCH8) Conscience de l'utilisateur

Résultats des études

Les expériences passées, les attentes, les aptitudes et les craintes des utilisateurs, sont autant d'éléments essentiels à l'élaboration d'un système adapté à ceux-ci.

Les expériences avec d'autres interfaces vocales et les connaissances techniques de l'utilisateur ont un impact significatif sur son modèle mental de l'interaction vocale. Selon Corbett *et*

al.[14], les utilisateurs adoptent le modèle mental d'autres agents conversationnels populaires lorsqu'ils utilisent un nouveau système. Cette observation est consistante avec celle de Myers *et al.*[47] dont les résultats indiquent l'importance des biais d'assimilation lors de l'utilisation d'une nouvelle interface vocale. À l'inverse, ils retiennent qu'une expérience en programmation n'a pas d'impact généralisé sur les performances des utilisateurs mais contribuent à ce que ceux-ci adoptent un style injonctif pour interagir avec le système. Luger *et al.*[41] ont constaté que les compétences techniques des participants influencent positivement leur modèle mental. Leurs attentes concernant les capacités et l'intelligence du système sont plus modestes. Ils ont tendance à pardonner plus facilement les erreurs, à en localiser les causes au sein de la technologie, et à persévérer jusqu'à accomplir la tâche désirée. Si les connaissances techniques n'améliorent pas forcément les performances, elles contribuent à établir un modèle mental se rapprochant des capacités du système. Ainsi, Myers *et al.*[48] ont observé qu'un modèle mental erroné favorise la tentative d'exécution d'intentions non-supportées par le système. Certaines caractéristiques cognitives telles que la mémoire à court-terme favorisent une interaction compréhensible par le système. Wolters *et al.*[83] ont observé qu'au plus la mémorisation des chiffres (digit span) et l'affinité à la technologie sont élevées, au plus les utilisateurs utilisent de mots-clés et partagent un vocabulaire commun avec le système. Ces mêmes utilisateurs sont plus enclins à noter l'interaction positivement et à en avoir une impression d'ensemble favorable.

Les préférences et attentes de l'utilisateur permettent de distinguer les fonctionnalités et comportements du système jugés indispensables par les utilisateurs. Indépendamment de la forme d'interaction la plus appropriée au contexte d'utilisation, les utilisateurs désirent entretenir un style conversationnel avec le système selon Corbett *et al.*[14]. Les participants de l'étude de Portet *et al.*[57] préfèrent converser avec l'habitat et non avec une entité matérialisée tel qu'un robot même si seulement la moitié des participants trouvent naturel d'interagir par la parole. Callejas *et al.*[11] ont demandé à leurs participants quels étaient les aspects les plus importants à leurs yeux entre : fonctionnement correct, amabilité de la synthèse vocale, et facilité d'utilisation. La plupart ont placé l'amabilité du système et la facilité d'utilisation avant le fonctionnement correct. Selon Myers *et al.*[47], les utilisateurs veulent avoir l'initiative dans le dialogue. Les utilisateurs de l'étude de Pradhan *et al.*[58] ont exprimé le désir de pouvoir interagir avec plusieurs modalités tout en gardant la parole comme modalité principale. Enfin, Cowan *et al.*[15] mettent en évidence le souhait des utilisateurs d'avoir un système plus personnalisable et mieux intégré à son environnement.

Les craintes et les préoccupations des utilisateurs sont souvent des freins à l'adoption d'une technologie. Les personnes âgées de l'étude de Vacher *et al.*[75] ont la crainte de devenir dépendantes et de ne pouvoir agir seules alors que les personnes malvoyantes ont peur de perdre le contrôle de leur habitat. Selon Portet *et al.*[57], les utilisateurs âgés pensent qu'un environnement intelligent est dédié aux personnes ayant divers troubles. Leur orgueil prend le dessus, ils ne veulent pas donner l'impression d'être dépendants ou assistés en adoptant un tel système. Ils ont également peur de perdre en autonomie. L'inactivité est citée comme un risque de réduction de leurs capacités intellectuelles par manque de pratique. Les plus jeunes participants de l'interview de Callejas *et al.*[11] partagent cette dernière crainte. Pour eux, laisser le système agir à leur place entraîne une perte de compétences plus rapide. D'après Cowan *et al.*[15] et Abdorahmani *et al.*[1], la gêne sociale ressentie en utilisant une interface vocale dans les milieux publics est un obstacle cité par les utilisateurs pour ne pas recourir à l'interaction vocale en public.

La connaissance des habitudes et aptitudes de l'utilisateur est cruciale afin de concevoir des systèmes en accord avec leurs facultés. Parmi les vingt-deux études, seules deux présentent un système adaptif : le projet ALADIN[18, 25]. Malheureusement, les évaluations des utilisateurs de ces deux études ne portent pas sur la capacité du système à s'adapter aux facteurs linguistiques et acoustiques de l'interaction. Néanmoins, le manque d'adaptation aux aptitudes de

l'utilisateur est signalé par les utilisateurs des études de Pradhan *et al.*[58] et Abdolrahmani *et al.*[1]. Certains participants rapportent des problèmes d'accessibilité des repères visuels utilisés par les appareils connectés. L'absence de moyens alternatifs pour communiquer ces fonctionnalités peut s'avérer frustrant pour les personnes malvoyantes.

Interprétation des résultats

Caractéristiques de l'utilisateur : Les caractéristiques de l'utilisateur telles que ses connaissances, ses expériences passées, sa mémoire à court-terme, ses préférences, ses besoins, ses attentes, ses aptitudes et ses affinités avec la technologie influencent son expérience de l'interaction vocale. Connaître l'utilisateur est donc essentiel pour lui assurer une expérience efficace, satisfaisante et plaisante.

Attentes de l'utilisateur et intelligence du système : Les connaissances techniques et les expériences avec d'autres interfaces vocales contribuent, entre autres, à fixer les attentes initiales de l'utilisateur avec une nouvelle interface[47, 14]. Les attentes peuvent ne pas être en accord avec ce que propose le système et nuire à l'expérience de l'utilisateur. Les qualités anthropomorphiques de l'interface visant à rendre son usage agréable telles que les interactions humoristiques amènent les utilisateurs à créer un modèle mental de l'interaction vocale erroné[41]. Les personnes techniquement compétentes sont plus capables de voir au-delà des qualités artificielles pour concevoir leurs propres modèles mentaux d'interaction. A contrario, les personnes dont le bagage technique est plus modeste ont des attentes dépassant les capacités du système. Lorsque leurs attentes ne sont pas rencontrées, leur frustration les pousse à remettre en question l'intelligence du système. Une conception réfléchie de l'interaction vocale devrait aider à clarifier les capacités du système en révélant son intelligence. Les promesses interactives faites par l'engagement humoristique devraient être reconsidérées afin que de tels engagements aident à l'évaluation de l'intelligence du système par l'utilisateur[41]. *Modèles d'utilisateur* : L'usage de modèles peut aider à distinguer différents types d'utilisateurs et adapter le système selon le type retenu[40]. Cela implique que l'utilisateur définisse à quel modèle il corresponde. Ces modèles peuvent être départagés selon le niveau d'expérience de l'utilisateur relatif à l'interaction vocale, son affinité avec la technologie, ou toute autre information jugée pertinente. Outre l'usage de modèles génériques, un profil d'utilisateur peut être complété et utilisé par le système pour choisir la stratégie de gestion du dialogue la plus adaptée aux caractéristiques de l'utilisateur.

Adaptation à l'utilisateur : Les interfaces adaptatives sont conçues pour adapter leur comportement aux besoins de l'utilisateur[24]. L'adaptation est inférée depuis le contexte d'utilisation, les performances et les habitudes de l'utilisateur. Nombre de paramètres peuvent être ainsi traités pour en déduire le comportement idéal de l'interface. Le temps de réponse de l'utilisateur, la vitesse d'exécution, le nombre d'erreurs commises, la répétition de certaines actions, la localisation de l'utilisateur sont des exemples de paramètres enregistrés par le système. Par exemple, il est possible de limiter le retour aux seules informations pertinentes au-delà des premières utilisations ou d'adapter le contenu et les messages du système à la fréquence d'utilisation. L'utilisation de déclencheurs pour adapter dynamiquement l'interface ne doit pas se faire au détriment du sentiment de contrôle de l'utilisateur. Certains déclencheurs ne reflètent pas nécessairement les compétences de l'utilisateur. Prévoir la possibilité de désactiver ou de personnaliser cette fonctionnalité doit donc être prévu lors de la conception du système.

Aptitudes de l'utilisateur : Une meilleure compréhension des besoins des utilisateurs passe par une connaissance de leurs aptitudes. La conception d'interfaces doit se faire en tenant compte de l'éventail des aptitudes des futurs utilisateurs. Le choix d'une modalité, d'un comportement ou d'une fonctionnalité du système doit se faire en veillant à proposer une solution alternative accessible pour les populations à besoins spécifiques. Une conception participative peut

constituer un moyen d'aider à concevoir des systèmes en accord avec leurs besoins.

Autonomie et dépendance : Les populations à besoins spécifiques ont clairement montré qu'elles étaient très prudentes en n'acceptant pas les systèmes qui les pousseraient dans une situation de dépendance[57]. Elles veulent garder le contrôle et ne pas voir leur autonomie réduite plutôt que d'être améliorée. Elles veulent continuer autant que possible à effectuer des activités nécessitant des efforts physiques et cognitifs afin que leur condition ne se détériore pas précipitamment. Le système doit être limité à l'assistance des populations à besoins spécifiques sans forcément agir à leur place. Il est important de ne pas leur donner le sentiment d'être des personnes assistées pour l'ensemble de leurs tâches quotidiennes.

3.3.9 (VOCH9) Conscience du contexte d'utilisation

Résultats des études

Le contexte d'utilisation fait référence à l'ensemble des circonstances, des événements, ayant un rapport direct ou non à l'interaction entre l'utilisateur et le système. Parmi les notions indirectes à l'interaction, il y a la localisation de l'utilisateur, son occupation actuelle, l'environnement dans lequel il évolue ou le moment de la journée. Le contexte d'utilisation peut aussi se rapporter à des données inhérentes à l'échange entre l'utilisateur et le système telles que l'ensemble des requêtes ayant précédées une commande particulière. La récolte et le traitement de ces informations ont pour but d'améliorer l'expérience de l'utilisateur.

L'environnement entourant l'utilisateur peut perturber son interaction avec le système. Dans les milieux publics, plusieurs utilisateurs de différentes études se sentent embarrassés d'interagir à voix haute avec une intelligence artificielle. C'est le cas pour les participants des études de Cowan *et al.*[15] et Abdolrahmani *et al.*[1]. Dans cette dernière étude, les personnes malvoyantes préfèrent éviter l'interaction de peur que des informations sensibles ou personnelles soient perceptibles par des inconnus. Ils suggèrent différentes solutions telles que l'utilisation d'un casque-micro ou bien que le système s'exprime plus rapidement. Selon eux, un débit plus élevé est inintelligible pour la plupart des personnes n'ayant pas de troubles visuels sévères. À l'inverse, ils considèrent leur domicile comme une zone de confort où ils peuvent s'exprimer librement.

La sensibilité des microphones des système à destination de l'habitat connecté pose question. D'après Abdolrahmani *et al.*[1], les utilisateurs ont relevé une activation inopinée du système lorsque les microphones sont placés trop près d'une télévision allumée. Cette sensibilité les pousse à couper les micros lors de rassemblements sociaux. Ils soulignent le manque de précision de la reconnaissance automatique de la parole dans les milieux bruyants, les sons ambiants masquent ou interfèrent avec les commandes énoncées verbalement. Les participants de l'étude de Cowan *et al.*[15] ont partagé les mêmes reproches.

À la manière d'une conversation entre êtres humains, les interlocuteurs montrent parfois un comportement laissant supposer que le système a une connaissance du contexte de leur échange. Lors d'une incompréhension du système, les participants tentent de corriger la commande prononcée en ajoutant des informations supplémentaires sans répéter l'entièreté de la commande. Ce comportement observé par Derboven *et al.*[18] traduit des attentes élevées des utilisateurs quant à l'intelligence du système. Celui-ci doit être capable de raisonner sur base des précédentes interactions. De même, en cas de commande imprécise ou incomplète, il doit pouvoir inférer du contexte de l'interaction afin d'accomplir l'action attendue. Selon Luger *et al.*[41], les utilisateurs font part d'une expérience plus positive lorsqu'ils perçoivent le système comme étant capable de prendre en compte les actions antérieures.

Interprétation des résultats

Contexte et intimité : La connaissance du contexte d'utilisation peut avoir des avantages tels que la possibilité d'améliorer la compréhension des requêtes, et de leurs intentions, par le système ou bien fournir un comportement en adéquation avec son environnement. Pour les appareils mobiles intégrant une interface vocale, une réaction appropriée du système est idéalement requise en fonction de la situation afin de préserver l'intimité de l'utilisateur. Dans certains contextes, le fait de fournir des modes «silencieux» ou «discret», dans lequel les détails personnels sont soit présentés de manière plus discrète, soit masqués, peut fournir la confidentialité nécessaire aux utilisateurs[1]. La collecte d'informations pertinentes, au travers d'une analyse continue des discussions contextuelles, pourrait améliorer la prise de décision du système[55]. Une telle solution pose la question du respect de l'intimité tant il est difficile de savoir ce que le système enregistre comme informations et ce qu'il en fait. Sans en arriver jusqu'à telle extrémité, le système pourrait se baser sur les requêtes antérieures ayant échouées afin d'améliorer la précision de la compréhension des intentions de l'interlocuteur lors de requêtes successives. Pour les utilisateurs soucieux de l'activation inopinée de l'interaction vocale lors de la reconnaissance d'un mot d'éveil dans un milieu bruyant, il doit être possible de désactiver les microphones du système aisément. Des modes dégradés, dans lesquelles le système n'est activable que par contact physique ou ne rassemble aucune donnée contextuelle, peuvent être aussi envisagés.

3.3.10 (VOCH10) Respect de l'intimité de l'utilisateur et amélioration de sa sécurité

Résultats des études

L'interaction vocale nécessite l'enregistrement et le traitement de commandes formulées verbalement afin d'exécuter les actions attendues. Le système doit donc être à l'écoute de l'utilisateur et est susceptible d'enregistrer des données privées de celui-ci. Bien qu'une telle capacité puisse lever des préoccupations relatives au respect de l'intimité des usagers, les participants de la majorité des études abordant ces questions en minimisent l'importance comme l'illustrent Kowalski *et al.*[36]. Pour Abdolrahmani *et al.*[1], les utilisateurs sont généralement prêts à accepter la menace de vol ou d'un usage non-désiré de leurs données comme un compromis raisonnable aux avantages qu'offre l'interaction vocale. En outre, ils indiquent que ces informations sont déjà exposées au travers d'autres services présentant des failles de sécurité. Si leur intimité peut être violée de la sorte, ils n'acceptent pas pour autant que des étrangers puissent entendre ces mêmes informations dans les transports publics. Les utilisateurs de l'étude de Portet *et al.*[57] ne considèrent pas leur intimité menacée par la technologie vocale même s'ils la trouvent trop intrusive dans leur vie privée. Peu de participants expriment des craintes concernant l'enregistrement de leurs conversations. A contrario, ils ne désirent pas être filmés dans leur domicile. Vacher *et al.*[75] nuancent un peu plus ces propos. Si les utilisateurs acceptent que leurs conversations soient enregistrées, ils ne veulent pas que ces enregistrements soient accessibles en dehors de leur habitat. La possibilité d'arrêter l'enregistrement à tout moment doit être prévue par le système. En contradiction avec les précédentes études, Pradhan *et al.*[58] mentionnent des inquiétudes des utilisateurs quant à l'écoute continue, l'enregistrement des interactions, la collecte de données personnelles et la probabilité de piratage des services en ligne. Les plus extrêmes vont jusqu'à désactiver les microphones ou débrancher leurs appareils connectés.

La sécurité physique des utilisateurs n'est que très peu traitée parmi les vingt-deux études. Pourtant, Zeng *et al.*[86] soulignent une préoccupation plus importante des utilisateurs pour les menaces physiques par rapport à la violation de leur intimité. D'une part, ils redoutent l'exploitation de failles des appareils critiques tels que des verrous intelligents. D'autre part, ils

sont enclins à utiliser des caméras et des éclairages programmés afin de dissuader les personnes malveillantes de pénétrer leur domicile. Dans un autre registre, les mécanismes de prévention telles que le signalement de portes non-verrouillées ou de gazinière allumée par des signaux sonores sont largement appréciés selon Portet *et al.*[57].

Zeng *et al.*[86] ont observé une corrélation entre un modèle mental sophistiqué des environnements intelligents et la connaissance de modèles de menace avancés. Les participants dont le modèle mental est le plus complet sont capables d'identifier les risques uniques relatifs aux environnements intelligents et de prendre des précautions spécifiques pour traiter ces risques. Inversement, un modèle mental incomplet implique un manque de prise en charge des différentes menaces et des comportements adéquats à adopter.

Interprétation des résultats

Environnements connectés et utilisateurs multiples : Les questions de confidentialité et d'intimité ne concernent pas seulement les utilisateurs principaux. Ceux-ci sont amenés à avoir des activités sociales avec d'autres individus. En pénétrant dans leur environnement, ces personnes sont susceptibles d'être entendues par le système[57]. Le système doit non seulement améliorer la sécurité de l'utilisateur mais également celle de leur entourage en contact avec le système.

Dans le cadre d'interactions multi-utilisateurs, le déséquilibre entre l'utilisateur principal et les utilisateurs secondaires en matière d'affinités avec la technologie vocale et de contrôle sur celle-ci doit également être pris en compte[86]. Il n'est pas concevable que le système délaisse les utilisateurs incidents et néglige leurs préoccupations en matière de sécurité.

Retour d'information et intimité : Améliorer les modèles mentaux des utilisateurs a un impact sur leurs modèles de menaces. La connaissance de ces modèles de menace confère aux utilisateurs la capacité à déterminer s'ils doivent ou non prendre action face aux risques pour leur sécurité ou leur vie privée. Un retour d'information adéquat sur l'activité des appareils connectés, comme par exemple des témoins lumineux semblables à ceux des caméras, permet d'attirer leur attention sur le fait que leurs conversations sont susceptibles d'être enregistrées[86]. Afin de prévenir l'aspect intrusif d'une interface vocale dotée de la reconnaissance automatique de la parole, le système devrait émettre un son bref pour avertir les utilisateurs avant d'entamer un dialogue avec eux[57].

Enseignement de bonnes pratiques : La communication avec les utilisateurs finaux est essentielle. Leur enseigner un ensemble de bonnes pratiques en matière de politique de sécurité permet de minimiser les risques et d'ouvrir leur conscience aux risques potentiels. Par exemple, désactiver les microphones lors de rassemblements sociaux ou lorsqu'ils ne sont pas jugés utiles ou encore avertir les invités que leurs conversations peuvent être enregistrées[86].

Compromis entre sécurité et fonctionnalité : En matière de sécurité et de confidentialité, le compromis est difficilement acceptable avec les fonctionnalités du système. Par exemple, ne pas forcer l'utilisation de services en ligne est un choix technique permettant de réduire les risques liés à la perte de connexion réseau ou à leur mise hors service[86].

3.3.11 (VOCH11) Entretien d'une relation de confiance basée sur la fiabilité et la transparence

Résultats des études

Il peut être difficile pour une technologie de gagner la confiance de ses utilisateurs. Le manque de transparence et de fiabilité conduisent les usagers à remettre en cause cette confiance et leur font adopter des comportements prudents. Selon Abdolrahmani *et al.*[1], les participants de leur étude sont frustrés par les actions inattendues du système dues à une incompréhension d'une commande vocale. Cette frustration détériore la confiance dans le système et implique

le développement de stratégies pour s'assurer de l'absence d'erreurs commises. Pour les tâches complexes, en particulier, les utilisateurs jugent plus négativement les assistants personnels des appareils mobiles. Ces assistants intelligents ont tendance à afficher une recherche en ligne contrairement aux solutions destinées à l'habitat dépourvues d'écran. Les participants de l'étude de Luger *et al.*[41] vont jusqu'à se limiter à des tâches basiques par manque de confiance en la capacité des agents conversationnels à traiter des actions complexes ou des données sensibles. Une inconsistance dans l'exécution des tâches même basiques, telles que la programmation d'une alarme ou un appel, entraîne une méfiance des utilisateurs comme l'ont remarqué Cowan *et al.*[15].

La résilience du système est l'une des préoccupations des utilisateurs comme le soulignent Zeng *et al.*[86]. Dans le cas d'une perte de connexion à internet, les utilisateurs veulent que leurs appareils connectés continuent de fonctionner normalement. Les appareils liés à la domotique nécessitant une alimentation électrique doivent également être contrôlable lors d'une panne d'électricité.

La plupart des solutions vocales à destination du grand public sont le fruit de grandes compagnies du monde de l'informatique. Au-delà de la confiance dans un système particulier, c'est le manque de transparence de la compagnie proposant cette solution qui est remise en question. D'après Cowan *et al.*[15], les participants de leur étude sont soucieux de l'usage de leurs données sensibles. Celles-ci peuvent être collectées et monétisées par les grandes compagnies à leur insu.

Interprétation des résultats

Confiance dans les actions du système : Certaines requêtes peuvent être plus sensibles que d'autres. L'envoi de courriels ou de messages à caractère professionnel, l'enregistrement d'un rendez-vous important dans l'agenda, ou encore le verrouillage de la porte d'entrée en sont des exemples parmi d'autres. Pour toutes ces actions, les utilisateurs sont amenés à investir du temps et des efforts pour vérifier la bonne exécution de leurs commandes. Pour maintenir la confiance de l'utilisateur dans les actions entreprises, des phrases de confirmation peuvent être incluses au dialogue pour donner la certitude à l'utilisateur que ses intentions ont correctement été traduites et exécutées par le système[1, 85].

Tolérance aux pannes : La dépendance de l'interaction vocale et des appareils connectés à une connexion internet ou au réseau électrique ne doit pas empêcher leur bon fonctionnement. En toutes circonstances, l'utilisateur doit être en mesure d'interagir avec son environnement quitte à basculer vers une interaction physique. La dépendance au réseau internet doit être limitée autant que possible pour permettre au système de fonctionner même sans connexion.

Transparence de l'usage des données : Potentiellement, l'interaction vocale permet d'enregistrer une grande quantités d'informations sensibles et personnelles à l'insu de l'utilisateur. Il n'est pas tolérable que les données des usagers soient utilisées dans d'autres buts que d'améliorer leur propre expérience du système. Dans le cas où ces données sont collectées, les fournisseurs de solutions basées sur l'interaction vocale doivent déclarer clairement et précisément l'usage et la portée des informations de l'utilisateur afin qu'il soit conscient des possibles violations de la confidentialité de ses discussions[15]. La confiance des utilisateurs peut être aussi renforcée en leur donnant accès aux données enregistrées et la possibilité de choisir la manière dont elles seront exploitées.

3.3.12 (VOCH12) Correspondance avec le monde réel

Résultats des études

L'interaction vocale tend à prendre une place de plus en plus prépondérante dans la manière de gérer son environnement et organiser sa vie au quotidien. En particulier, Les systèmes conversationnels imitent les échanges humains afin de proposer une expérience agréable et naturelle aux utilisateurs. Luger *et al.*[41] ont observé une tendance des utilisateurs à parler à l'interface vocale comme s'il s'agissait d'un être humain. Lorsqu'ils se retrouvent seuls dans un environnement contrôlable par la parole, les participants de l'étude de Vacher *et al.*[75] recherchent un interlocuteur à qui s'adresser. De même, lorsque le système répond à leurs demandes, ils se demandent vers qui ou quoi se tourner. Les personnes malvoyantes de cette même étude réagissent différemment, ils se concentrent uniquement sur l'exécution de leur commande. Le sentiment de solitude varie également entre les deux populations. Là où la population malvoyante conserve un sentiment de solitude, les personnes âgées ont l'impression d'avoir une présence avec eux. Matérialiser cette présence est délicat comme le font remarquer Callejas *et al.*[11]. Plusieurs interviewés ne veulent pas que le système revête une apparence humaine, même si cette apparence est un personnage d'animation. Par contre, Ils ne voient pas d'objection à utiliser le visage d'un proche. Ces avis concordent avec ceux exprimés dans l'étude de Portet *et al.*[57]. Les utilisateurs ne désirent pas converser avec un robot ou un point de contrôle à distance mais sont unanimes sur le fait d'utiliser une voix naturelle. Cowan *et al.*[15] soulignent l'impression positive des utilisateurs à l'écoute d'une voix synthétisée aux caractéristiques humaines. Ils apprécient également les nuances culturelles apportées à la voix et à la personnalité du système. Cependant, ils émettent quelques doutes sur l'efficacité d'une telle métaphore dans certains contextes particuliers.

Interprétation des résultats

Anthropomorphisme de l'interface vocale : Les utilisateurs ont tendance à attribuer des qualités anthropomorphiques à l'interface vocale au travers des dialogues. Quand bien même ces dialogues sont exclusivement à l'initiative de l'utilisateur et que le système n'y réponde que par de brefs messages manquant de naturel, il semblerait qu'en utilisant la parole l'interlocuteur se met spontanément dans une situation de dialogue comme s'il s'agissait d'une conversation entre êtres humains[75]. L'interaction vocale ne doit pas laisser l'utilisateur présumer de son intelligence sur des impressions trompeuses. Il convient d'adopter un style de dialogue adéquat pour faciliter la création d'un modèle mental plus proche des capacités réelles et du niveau de complexité des tâches menées par le système dans le but d'améliorer l'expérience de l'utilisateur[15]. Même si le désir d'une interaction plus conversationnelle est exprimé, l'humanité n'est pas toujours une métaphore adéquate pour certaines tâches spécifiques tant elle crée une inadéquation entre les attentes des utilisateurs et les performances du système.

L'anthropomorphisme a ses limites. Un sentiment de malaise peut apparaître lorsque le système arbore une attitude ou une représentation trop humaine. Le réalisme nuit à son adoption et provoque une sensation désagréable chez les utilisateurs. Il est conseillé de restreindre les qualités humaines du système pour éviter toute forme de rejet.

3.3.13 (VOCH13) Support de l'intégration

Résultats des études

Au sein des environnements intelligents, plusieurs technologies différentes sont amenées à cohabiter. L'interaction vocale, l'une des principales modalités d'interaction de ces environnements, y joue un rôle important. Contrôler l'ensemble des appareils connectés via une interface

vocale unique n'est pas toujours garanti. Pour Cowan *et al.*[15], leurs participants voient l'intégration entre les appareils comme un moyen d'améliorer l'expérience utilisateur. Les utilisateurs de l'étude de Zeng *et al.*[86] abondent dans ce sens : ils veulent plus d'interopérabilité entre les diverses solutions existantes. Le manque de compatibilité et donc de diversité est d'ailleurs une crainte qu'expriment les utilisateurs comme l'ont découvert Kowalski *et al.*[36].

Pour certaines populations, une meilleure intégration contribue à une plus grande sensation d'indépendance. Elles peuvent contrôler un plus grand nombre d'appareils qui leur étaient autrefois inaccessibles. Ainsi Abdolrahmani *et al.*[1] arguent que cet avantage offre des opportunités intéressantes pour les populations malvoyantes. Les utilisateurs de leur étude accomplissent leurs tâches plus rapidement et font moins d'efforts pour interagir avec des solutions ou des applications tierces. Cependant, des améliorations sont encore nécessaires : quelques utilisateurs désirent que le système supporte l'usage de la parole pour un plus large éventail d'applications et d'appareils.

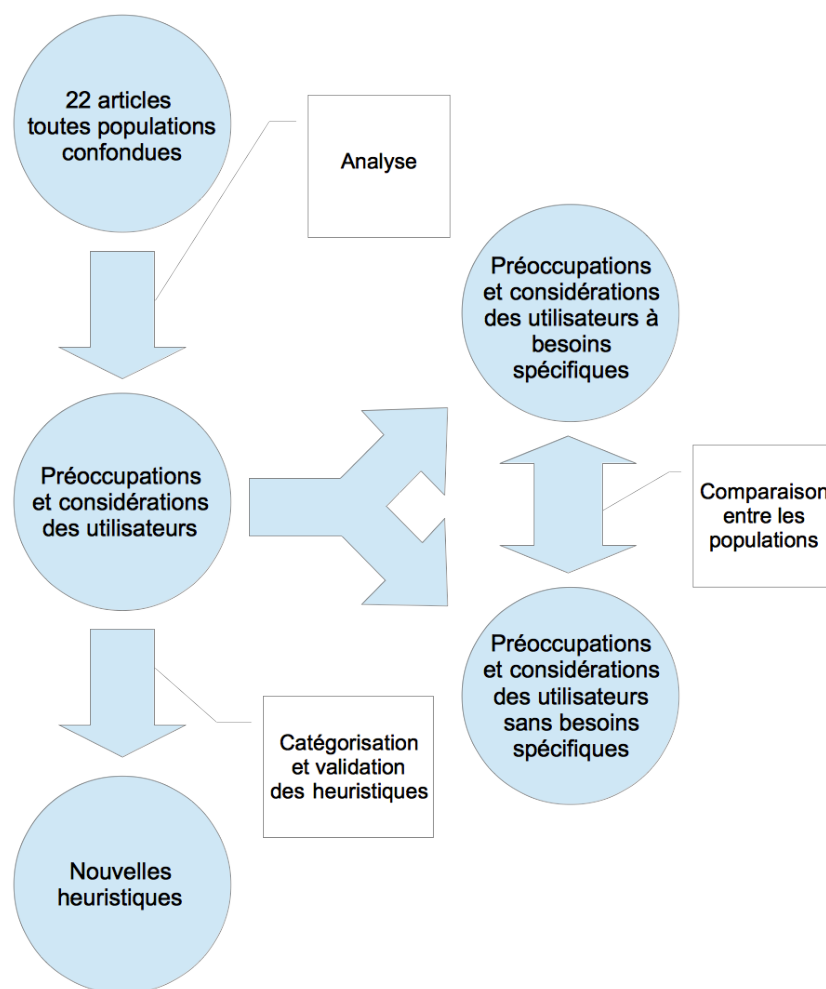
Interprétation des résultats

Centralisation et interopérabilité : Pour différentes populations à besoins spécifiques, et notamment les personnes malvoyantes, le contrôle centralisé des dispositifs connectés de leur habitat par la parole est parfois la seule forme de technologie leur permettant de rester indépendantes[1]. De nombreux utilisateurs ont recours à des appareils connectés fonctionnant avec des protocoles différents ou utilisent plusieurs applications avec des interfaces vocales concurrentes. Il est important de veiller à une interopérabilité entre toutes ces solutions la plus complète possible[15]. L'adoption de plusieurs protocoles par le dispositif central est l'une des pistes à explorer. L'intégration de tous ces composants provoque des apparitions de failles de sécurité aux bordures de chaque composant, veiller à une interopérabilité robuste est une nécessité pour s'assurer d'un bon fonctionnement du système dans son ensemble[86].

3.4 Évaluation par heuristiques : Analyse et discussion

Résumé

La catégorisation de l'ensemble des préoccupations des utilisateurs a confirmé la pertinence et l'intérêt de nouvelles heuristiques. Cependant, en comparant les résultats des analyses, il apparaît que des disparités existent entre les populations à besoins spécifiques et les autres populations. De ce fait, seules les préoccupations des populations à besoins spécifiques sont retenues pour l'élaboration des recommandations de conception. Une analyse plus minutieuse des treize articles ciblant ces populations a été réalisée et leurs préoccupations ont été ajoutées en annexe.



Graphique 3.2 – Validation des heuristiques et comparaison entre les populations

3.4.1 Validation des heuristiques

En se basant sur des travaux existants et sur un ensemble de caractéristiques définies suite à une étude exploratoire, treize heuristiques ont été élaborées ou adaptées pour correspondre au contexte de l'interaction vocale. En analysant les résultats de vingt-deux articles différents, il apparaît que ces heuristiques ont permis de catégoriser la totalité des points abordés dans les évaluations des utilisateurs ainsi que les discussions et recommandations afférentes. Les treize heuristiques sont donc susceptibles de regrouper et classer les problématiques, préoccupations et attentes spécifiques à l'interaction vocale mais également les solutions envisageables. La validation étant faite, il est nécessaire de vérifier si des recommandations de conception pour les principaux intérêts de chaque population selon les différentes heuristiques sont nécessaires ou non.

3.4.2 Comparaison entre les différentes populations

À la lumière des résultats des évaluations, il est possible de vérifier l'hypothèse selon laquelle les populations à besoins spécifiques et les autres populations⁵ rencontrent les mêmes problématiques et expriment les mêmes préoccupations. Étant donné la diversité des systèmes employés pour les différentes expérimentations, il est peu probant de comparer les résultats sur

5. Par la suite, ces populations seront qualifiées de populations ordinaires, moyennes, communes, lambdas, et ce, sans la moindre connotation péjorative, par opposition aux populations plus particulières que sont les populations à besoins spécifiques

la seule base de l'interaction vocale. Parmi les études, la plus large représentation est celle des agents conversationnels des géants du Web, ou GAFAM⁶, dont les variations en matière de fonctionnalités et d'utilisabilité sont moins marquées. Quatre d'entre eux proposent des agents personnels intelligents aux caractéristiques conversationnelles (Alexa d'Amazon, Siri d'Apple, Cortana de Microsoft et l'Assistant de Google) afin de fournir tout un ensemble de services aux particuliers allant de la possibilité de régler une alarme jusqu'au contrôle de son environnement par le seul usage de la parole. Un tiers des articles, soit sept articles, reposent sur une expérimentation avec l'un de ces agents conversationnels. Trois études ont pour cible des populations ordinaires[15, 41, 55], trois autres études s'intéressent aux personnes malvoyantes[1, 58, 4] et une dernière étude se focalise sur les personnes âgées[36].

Les populations lambdas partagent des inquiétudes et des problématiques relatives au respect de l'intimité et de la confidentialité de leurs données avec les autres populations. Ils se posent des questions quant à la collecte et l'usage de leurs données privées par les grandes compagnies[15]. Les personnes malvoyantes sont également préoccupées par la collecte de leurs informations et le fait que le système soit toujours à l'écoute de leurs conversations[58]. Interagir vocalement dans les milieux publics induit une sensation de malaise, de gêne sociale, tant pour les personnes ordinaires[15] que les personnes atteintes de cécité[1]. Cela les pousse à ne pas s'engager dans un échange conversationnel avec le système en public[41, 1]. D'ailleurs, les deux populations mentionnent des incompréhensions de la reconnaissance automatique de la parole dans les milieux bruyants[15, 1]. Les faiblesses du système nuisent à la confiance que lui portent les utilisateurs moyens[41, 15] et malvoyants[1]. Ils se limitent pour la plupart à des tâches basiques de peur que les tâches plus complexes échouent.

Même s'ils ne l'expriment pas exactement de la même manière, les utilisateurs ordinaires et ceux ayant des troubles visuels font état de problèmes nuisant à l'accessibilité de la technologie. Tout d'abord le non-respect de l'interaction à main-libre empêche les utilisateurs communs d'utiliser le système en toutes circonstances[15]. Garantir une interaction entièrement vocale est profitable à toutes les populations indépendamment de leurs besoins. Les populations malvoyantes pointent un manque d'alternatives accessibles aux témoins lumineux des appareils connectés. Ces témoins véhiculent bien souvent une information pertinente que ces populations ne perçoivent pas[58, 1].

Plusieurs fonctionnalités sont désirées de manière égale entre toutes les populations. Un souhait récurrent est celui de pouvoir contrôler un nombre toujours plus élevé d'appareils au sein de l'habitat. Pour les personnes affectées par une perte de la perception visuelle tout comme les personnes âgées, une meilleure intégration et interopérabilité contribuent à gagner en indépendance et en efficacité[1, 58, 36]. Pour les autres populations, cela contribue à leur garantir une meilleure expérience de l'interaction vocale[15]. C'est d'ailleurs pour cette même raison que les utilisateurs lambdas désirent avoir la possibilité de personnifier l'interface vocale selon leurs préférences[15]. Les personnes malvoyantes font aussi état d'un besoin de pouvoir personnifier la synthèse vocale pour mieux correspondre à leurs aptitudes[1]. Une dernière fonctionnalité appréciée est la nature anthropomorphique de l'échange avec le système. Les personnes âgées et celles atteintes de cécité désirent converser avec une interface à la voix naturelle[36, 58]. Les avantages d'un échange conversationnel sont également appréciés par les populations communes[41] jusqu'à un certain point[15]. Ils remettent en question l'utilité d'une telle métaphore dans des contextes particuliers où la brièveté et l'efficacité priment.

Toutes les préoccupations ne sont pas pour autant partagées entre les différentes populations. Pour les populations à besoins spécifiques, l'interaction vocale est avant tout une technologie contribuant à les assister au quotidien. Elles sont donc plus critiques lorsque le système ne parvient pas à remplir ce rôle efficacement ou manque de considération pour leurs aptitudes. En particulier, les personnes malvoyantes trouvent l'édition de texte et la correction des erreurs

6. Acronyme de Google, Apple, Facebook, Amazon et Microsoft

fastidieuses[4, 1]. Elles déplorent un manque de flexibilité et des difficultés de contrôle sur l'interaction comme l'absence de modalités alternatives[58], des activations inopinées du système, et des limites dans la formulation des requêtes vocales[1]. Par contre, ces mêmes personnes ainsi que les personnes plus âgées[36] acceptent plus volontiers les entorses au respect de leur intimité si cela contribue à améliorer leur quotidien. Enfin, les personnes malvoyantes éprouvent des difficultés d'apprentissage des commandes[58].

Les populations ordinaires ont des attentes qui peuvent être en désaccord avec les capacités du système arguant d'une faiblesse de leur modèle mental. Ce modèle mental est tributaire de leurs connaissances et limite leur compréhension de l'interaction[41]. Ils ont recours à des réparations conversationnelles en cas d'erreur sans répéter l'entièreté de la commande vocale[41, 55]. En cas d'erreurs répétées, ils se cantonnent à des tâches basiques et considèrent la technologie vocale comme une distraction ludique.

En l'état, il n'est pas possible de confirmer l'hypothèse introduite plus tôt⁷ selon laquelle, quelque soit la population ciblée, les problématiques rencontrées sont les mêmes. Bon nombre des problèmes relevés par les populations à besoins spécifiques sont en réalité des préoccupations récurrentes au sein de l'ensemble des populations. Cependant, leurs différentes attentes et aptitudes fixent leur tolérance à certaines problématiques dont elles ne font pas toutes l'expérience. La difficulté d'édition en est un exemple. Les populations à besoins spécifiques sont sans doute celles qui profitent le plus des améliorations du système. Le gain en indépendance et en efficacité qu'elles en retirent contribue largement à faciliter l'accomplissement de leurs tâches quotidiennes. Toutes les améliorations allant dans ce sens sont bonnes à prendre et ne peuvent que profiter aux autres populations. Ces autres populations sont par contre plus difficiles à combler. Elles se désintéressent facilement de l'interaction vocale si elle ne rencontre pas leurs attentes[41]. Les satisfaire n'implique pas forcément une amélioration de l'expérience des autres populations. Les recommandations de conception sont donc plus pertinentes si elles visent les populations à besoins spécifiques. Cela ne peut que profiter à l'ensemble : chacun est amené à expérimenter des situations où ses aptitudes sont diminuées pour se rapprocher de l'expérience des personnes à besoins spécifiques[82]. Pour paraphraser les propos de Zajicek *et al.*[85]⁸, une bonne conception pour les personnes à besoins spécifiques est une bonne conception pour tous.

3.4.3 Préoccupations des populations à besoins spécifiques

Il convient d'identifier les principales préoccupations des différentes populations et en particulier des populations à besoins spécifiques en préambule de la rédaction des recommandations de conception. Même si les diverses populations partagent une grande partie de leurs préoccupations, il existe des différences quant à l'importance qui est accordée à chacune d'entre elles. Pour les personnes à besoins spécifiques, la prise en charge de leurs aptitudes et de leurs préférences ainsi que la facilité d'utilisation et l'efficacité jouent un rôle primordial dans l'acceptation de la technologie vocale. Pour les autres populations, les fonctionnalités, le fonctionnement correct, et leur compréhension du système, sont leurs principaux intérêts.

Les tableaux en annexe⁹ reprennent l'ensemble des préoccupations exprimées par les utilisateurs ou observées par les chercheurs dans le cadre d'une enquête ou d'une expérimentation avec une interface vocale. L'objectif étant de dégager une série de recommandations à destination des populations à besoins spécifiques, seules les treize études ciblant ce type de population sont considérées. Chaque point abordé est catégorisé selon une heuristique parmi les treize définies auparavant. Dans quelques cas, une préoccupation peut être mise en relation avec plus d'une heuristique.

7. Voir point 3.2.1

8. "Good design for older adults is good design for everybody"

9. Voir Annexe A

Il apparaît que les heuristiques les plus représentées sont celles relatives au contrôle du système, à la recherche d'efficacité et de flexibilité, ainsi qu'au respect de l'intimité des utilisateurs (voir tableau 3.5). La manière dont les utilisateurs interagissent avec l'interface vocale est généralement liée à une volonté d'efficacité. Les deux heuristiques sont souvent utilisées pour catégoriser les mêmes préoccupations. Les populations à besoins spécifiques ont recours à une large variété de styles et de moyens différents pour contrôler le système sans pour autant qu'une tendance se démarque plus des autres. Certains problèmes limitent le contrôle des utilisateurs sur l'interaction et renforcent l'impression d'interagir avec un système autonome. La présence de plusieurs modalités d'entrée et de sortie ainsi que la personnalisation de l'interaction selon leurs préférences contribuent à l'efficacité des utilisateurs mais également à la prise en charge de leurs aptitudes. Quant aux questions relatives au non-respect de leur intimité, elles mettent en lumière quelques divergences au sein des populations à besoins spécifiques même si elles sont enclines à considérer la violation de leur vie privée comme un compromis nécessaire pour une meilleure interaction.

À l'autre bout du spectre, on retrouve les heuristiques liées à la découverte des commandes vocales, à l'aide contextuelle, à la prévention des erreurs et à la confiance accordée au système. Le manque de considération pour les erreurs de compréhension de l'interface vocale peut trouver une explication dans l'absence de solutions alternatives simples et abordables aux problèmes d'accessibilité. Les populations à besoins spécifiques acceptent le manque de fiabilité du système tant les avantages qu'elles en tirent améliorent grandement leur autonomie. Le plus étonnant est le faible nombre de préoccupations concernant l'apprentissage et la découverte des commandes vocales. Une seule étude mentionne des difficultés à apprendre à interagir avec le système sans qu'aucune raison ne soit donnée expliquant ce fait. La multimodalité du système, la familiarité des utilisateurs avec l'interaction vocale et l'expérience avec d'autres interfaces similaires sont autant d'hypothèses permettant d'apporter des éléments de réponse. La multimodalité fournit une interaction plus riche et un retour significatif permettant aux utilisateurs de mieux appréhender le système. La familiarité des populations à besoins spécifiques avec les outils d'accessibilité et en particulier la technologie vocale peut contribuer à formuler des requêtes verbales de manière adéquate. Enfin, en extrapolant les problématiques relatives aux biais d'assimilation, on peut en déduire que l'expérience des utilisateurs avec d'autres interfaces vocales les aide à forger un modèle mental pas si éloigné du système expérimenté. La vérification de ces hypothèses n'entre pas dans le cadre de ce travail.

Heuristique	Préoccupation/considération	nombre
VOCH1	Considérations liées à la manière d’interagir avec le système	11
	Manque de contrôle sur le système	4
VOCH2	Inadéquation du retour d’information	4
	Personnalisation du retour d’information	4
VOCH3	Fréquence élevée du nombre d’erreurs	3
VOCH4	Répétition en cas d’incompréhension	2
	Difficultés liées à la récupération et la correction des erreurs	4
VOCH5	Difficultés d’apprentissage	1
	Pertinence de l’introduction contextuelle et des exemples introductifs	3
VOCH6	Pertinence de l’aide contextuelle	2
VOCH7	Considérations liées à l’efficacité	9
	Désir de multiples modalités d’interaction	2
VOCH8	Manque de prise en charge des aptitudes	2
	Sentiment d’être assisté ou dépendant	4
	Prise en charge de l’expérience de l’utilisateur	1
VOCH9	Manque d’intelligence contextuelle	5
	Préoccupations liées à l’interaction dans les milieux publics	1
VOCH10	Manque de prise en charge de la confidentialité, de l’intimité et de la sécurité des utilisateurs	3
	Acceptation des compromis impactant la confidentialité, la sécurité et l’intimité des utilisateurs	5
	Appréciation des solutions sécurisantes	1
VOCH11	Manque de confiance dans l’interaction et de fiabilité de celle-ci	4
VOCH12	Désir de réalisme dans l’interaction	2
	Malaise éprouvé par le réalisme de l’interaction	4
VOCH12	Gain en indépendance	3
	Difficultés de mise en oeuvre de l’intégration	2

Tableau 3.5 – Occurences des préoccupations et considérations des utilisateurs à besoins spécifiques

Chapitre 4

Design de dialogue pour populations à besoins spécifiques : Élaboration des recommandations de conception

4.1 Principes de la conception basée sur les aptitudes

L'identification des préoccupations des populations à besoins spécifiques a permis de soulever des problèmes d'accessibilité dans trois études sur les onze. Ces trois études[58, 1, 4] ont en point commun d'utiliser des solutions commerciales destinées au grand public. Les difficultés d'édition de texte[4], le manque d'intégration[1] et le manque d'accessibilité visuelle[58, 1] sont les principaux reproches formulés à l'encontre des appareils connectés (Google Home, Amazon Echo) et des interfaces vocales des acteurs majeurs du domaine de l'informatique. Il est intéressant de constater que les autres études, à l'exception d'une seule[36], ne concernent que des systèmes développés pour des personnes à besoins spécifiques. Les évaluations et observations ne font mention d'aucun problème d'accessibilité rencontré par les participants. Il existe donc un manque d'accessibilité des solutions qui ne sont pas spécialement conçues pour des populations spécifiques.

La rédaction des recommandations de conception a pour objectif de répondre aux besoins de populations à besoins spécifiques mais également de pallier les problèmes d'accessibilité. Pour ce faire, les recommandations appliquent les sept principes de la conception basée sur les aptitudes de Wobbrock *et al.*[82] (voir tableau 4.1). Ces principes sont divisés en trois catégories : l'attitude et l'approche du concepteur, l'adaptabilité de l'interface utilisateur, la faculté de détecter et de modéliser les performances des utilisateurs et leur contexte. Seule la première catégorie est indispensable, les deux autres étant optionnelles.

1. La position du concepteur : les concepteurs considèrent ce dont sont capables les utilisateurs et non l'inverse ou ce dont tout le monde est capable. Dans le processus itératif de conception, la charge du changement doit être imputée au système plutôt qu'à l'utilisateur. Dans le cas idéal, un système basé sur les aptitudes devrait être suffisamment flexible pour permettre aux utilisateurs d'utiliser le système sans avoir à modifier leur corps, connaissance ou comportement. Les concepteurs sont encouragés à utiliser, lorsque c'est possible, des périphériques et logiciels peu coûteux, directement disponibles et en vente libre tout en étant facilement remplaçables afin de réduire les barrières à l'accès.
2. L'interface adaptative ou adaptable : l'interface mérite une attention particulière car c'est sur elle que les aptitudes des utilisateurs sont exercées et produisent leurs effets. L'adaptabilité fait référence à la capacité des interfaces de s'auto-ajuster ou d'être ajustées en fonction des performances de l'utilisateur ou du contexte. La transparence assure aux utilisateurs la possibilité de contrôler les adaptations, en mettant l'emphasis sur le

fait de rester eux-mêmes plutôt que d'avoir à changer pour rencontrer les besoins de la technologie.

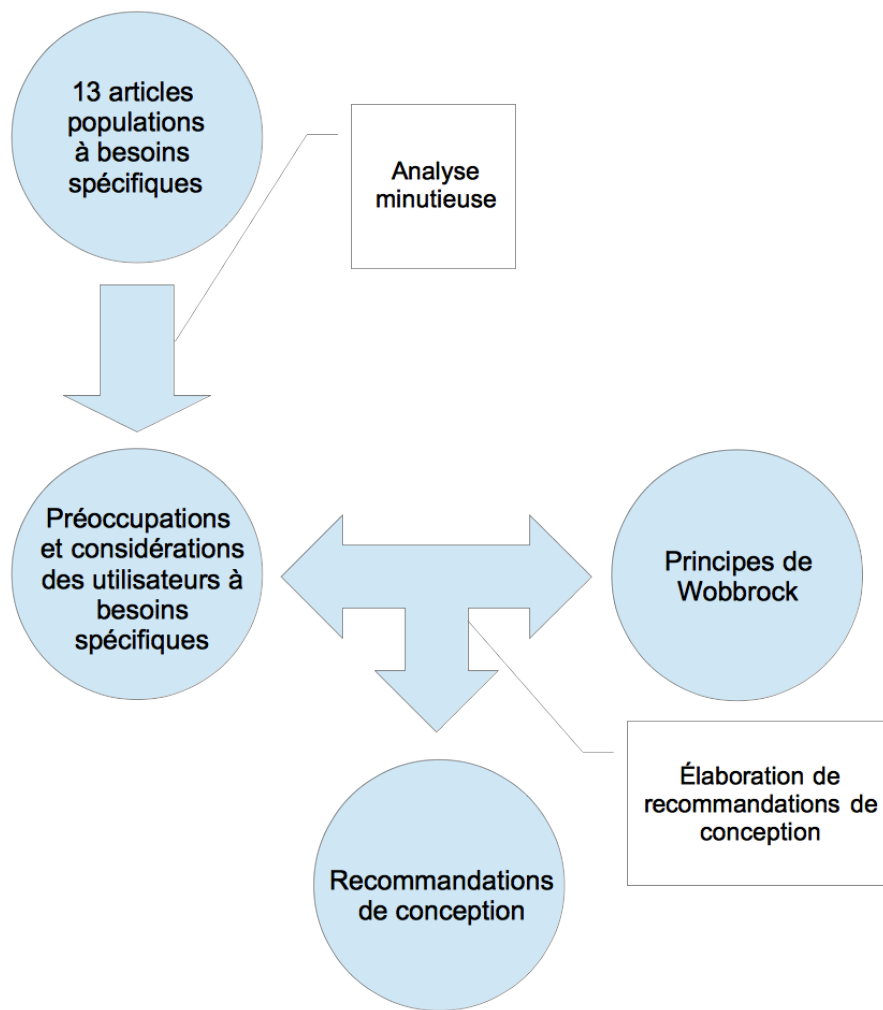
3. La détection et modélisation : les performances des utilisateurs et la connaissance du contexte permettent de tenir compte des actions des utilisateurs et de son environnement, éventuellement au moment de l'exécution, afin d'améliorer le système.

	Principe	Description
Position du concepteur (requis)	Aptitude	Les concepteurs se concentrent sur les capacités des utilisateurs, et non sur leurs handicaps, en s'efforçant d'exploiter tout ce dont les utilisateurs sont capables dans une situation, un contexte ou un environnement donné.
	Responsabilité	Les concepteurs répondent aux problèmes d'utilisabilité en modifiant les systèmes, et non les utilisateurs, les laissant tels qu'ils sont.
	Disponibilité	Les concepteurs utilisent des logiciels, matériels ou autres composants abordables et disponibles par des moyens accessibles.
Interface adaptative ou adaptable (optionnel)	Adaptabilité	Les interfaces peuvent être adaptatives ou adaptables pour fournir la meilleure adéquation possible aux aptitudes des utilisateurs.
	Transparence	Les interfaces peuvent informer les utilisateurs des comportements adaptatifs et de ce qui les régit, ainsi que des moyens d'inspecter, de contourner, d'éliminer, d'inverser, de stocker, de récupérer, de prévisualiser, de modifier ou de tester ces comportements.
Détection et modélisation (optionnel)	Performance	Les systèmes peuvent détecter, surveiller, mesurer, modéliser, afficher, prévoir ou utiliser autrement les performances des utilisateurs afin de fournir la meilleure adéquation possible entre les systèmes et les aptitudes des utilisateurs.
	Contexte	Les systèmes peuvent détecter, surveiller, mesurer, modéliser, afficher, prédire ou utiliser d'une autre manière la situation, le contexte ou l'environnement des utilisateurs pour anticiper les effets sur les capacités des utilisateurs et en tenir compte.

Tableau 4.1 – Sept principes de la conception basée sur les aptitudes[82]

4.2 Recommandations pour les populations à besoins spécifiques

Les trente recommandations suivantes répondent à la majorité des préoccupations exprimées par les utilisateurs dans les treize études retenues. Dans un souci de proposer une interaction accessible à un large panel de populations, les recommandations appliquent les principes de Wobbrock *et al.*[82] (voir tableau 4.1).



Graphique 4.1 – Création des recommandations de conception

Contrôle sur le système

- *Éveil du système à l'aide d'un mot ou d'une action spécifique* : Un système intégrant une reconnaissance automatique de la parole ne devrait pas être socialement interruptif, mais devrait pouvoir «se fondre» dans une interaction sociale normale. Sur cette base, l'adressage de l'interface vocale avec un nom, un mot clé ou en pressant un bouton est une exigence technique nécessaire pour distinguer la conversation naturelle des commandes vocales[18].
- *Dialogue avec le système* : En interagissant avec une interface vocale, les utilisateurs se mettent spontanément dans des situations de dialogue ou formulent le désir de converser avec le système. Afin d'éviter de contrarier les utilisateurs, l'interaction vocale devrait intégrer une gestion du dialogue même minimale[75].
- *Cohabitation de l'interaction physique avec les dispositifs connectés* : Le contrôle des utilisateurs sur leur environnement peut être assuré en conservant la possibilité d'interagir physiquement avec les appareils[86], et pas seulement via l'interface vocale. Cela peut également aider à atténuer les impacts potentiels des pannes de réseau, des pannes de service et jusqu'à un certain point, des pannes d'électricité. Un autre avantage est celui de permettre à certaines populations de continuer autant que possible à exécuter leurs tâches quotidiennes nécessitant des efforts physiques et cognitifs afin de rester en bonne condition. L'interaction vocale devrait pouvoir coexister avec les interrupteurs physiques et la manipulation directe des différents dispositifs.

Visibilité du système et retour d'information

- *Envoi de messages courts* : Un facteur important pour certaines populations, notamment les personnes âgées, est de limiter la taille des messages et instructions synthétisés par le système. Les messages devraient contenir au plus un élément d'information ou demander une seule donnée à la fois à l'utilisateur[85].
- *Personnalisation du retour d'information* : Le retour du système ne fournit pas toujours le niveau de précision ou de brièveté requis. Il devrait être possible pour l'utilisateur de personnaliser l'information délivrée par l'interface vocale en fonction de ses aptitudes, des ses préférences et du contexte d'utilisation[1]. Cela concerne aussi bien la qualité du retour que les moyens utilisés et le style adopté par la synthèse vocale.

Exemple : Certains utilisateurs sont incommodés par la verbosité du système. Une manière de pallier ce problème est d'avertir l'utilisateur par un bref passage sonore plutôt que par la synthèse vocale lorsque c'est possible.

- *Accessibilité visuelle* : L'information n'est pas toujours dispensée en prenant en compte les aptitudes des utilisateurs. Un retour visuel, tel que des signaux lumineux ou l'affichage sur un écran, peut manquer d'accessibilité pour certains besoins spécifiques. Le retour d'information devrait être accessible pour les personnes ayant des troubles visuels.

Exemple : Certains dispositifs du commerce arborent des voyants lumineux selon l'état du système (p. ex. : à l'écoute, en cours de configuration, défaillance du système) sans alternative accessible pour les malvoyants. Le système pourrait émettre un son caractéristique en fonction de ces différents états.

Prévention des erreurs

- *Adaptation de la syntaxe et du vocabulaire* : Certaines populations dévient des règles de langage utilisées pour interagir vocalement avec le système. La syntaxe et le vocabulaire de l'interaction vocale devraient pouvoir être adaptés dynamiquement, ou à la demande de l'utilisateur, afin de tenir compte des variations possibles tout en limitant par la même occasion les erreurs de compréhension.
- *Correction automatique par la connaissance du contexte* : En cas d'incompréhension d'une requête, le système devrait tenter d'inférer l'action attendue depuis les données contextuelles à sa disposition.

Exemples : Une écoute continue de la conversation pourrait améliorer la prise de décision du système en collectant des informations contextuelles pertinentes avant l'énonciation d'une requête[55]. D'autres manières de procéder seraient de raisonner sur l'ensemble des requêtes ayant échouées précédemment ou bien de déduire l'action à entreprendre selon les habitudes comportementales de l'utilisateur.

Identification et correction des erreurs

- *Identifier et renseigner clairement les erreurs* : Afin de faciliter la correction des erreurs, le système devrait exposer les motifs selon lesquelles la requête a échoué. De la sorte, les utilisateurs peuvent adopter la stratégie adéquate pour corriger l'erreur.
- *Adopter des mécanismes palliant les incompréhensions* : Les incompréhensions du système poussent les utilisateurs à répéter et articuler distinctement leurs commandes vocales sans y apporter d'autres modifications. La reconnaissance automatique de la parole n'est pas toujours en cause, il se peut que le système supporte la fonctionnalité désirée mais ne comprenne pas la syntaxe ou le vocabulaire utilisés. Dans ce cas, le système

devrait intégrer des mécanismes adaptés à l'utilisateur afin de l'aider à construire une requête correcte ou à clarifier les incompréhensions.

Exemples : Le système pourrait demander à l'utilisateur de formuler sa requête avec d'autres mots ou d'épeler un mot non reconnu. En s'inspirant des fonctionnalités de saisie automatique, l'utilisateur pourrait énoncer sa commande sans avoir à réfléchir sur des termes conduisant à une requête compréhensible[55].

- *Détecter les réparations conversationnelles :* Lorsque l'utilisateur se rend compte d'une erreur dans la formulation de sa requête, il a tendance à la compléter sans la répéter entièrement. À la manière d'une conversation naturelle, l'utilisateur suppose que son interlocuteur, en l'occurrence le système, connaisse le contexte de la correction. L'interface vocale devrait être à l'écoute de telles phrases de réparation et initier une séquence de réparation sur base de la requête précédente sans qu'elle ne soit répétée[55].
- *Éditer efficacement les textes :* L'édition par la parole d'un texte dicté s'avère être une entreprise tortueuse pour les utilisateurs au point qu'ils préfèrent revenir à des méthodes plus traditionnelles et moins accessibles. L'interaction vocale devrait inclure des fonctions d'édition de texte plus aisées et adaptées aux besoins des utilisateurs.

Exemple : Le système pourrait permettre de revenir sur n'importe quel mot et le corriger verbalement plutôt que de devoir répéter entièrement le texte dicté.

Aides et apprentissage des commandes

- *Apprentissage initial contextualisé :* L'apprentissage initial d'une interface vocale devrait introduire les utilisateurs à l'interaction avec le système en fournissant un bref aperçu des capacités du système avant de passer à un apprentissage dans un contexte d'utilisation[14].
- *Aide contextuelle :* Le matériel d'aide de l'interface vocale devrait être conçu autour d'actions disponibles et pertinentes dans un contexte spécifique[14]. L'aide devrait être intégrée au dialogue et non être exclusivement dépendante d'un autre support.
- *Invites d'aide adaptées :* Les invites d'aide ont un impact sur l'efficacité de certains utilisateurs. En fonction du profil de l'utilisateur et de ses performances, l'interface vocale devrait proposer des invites d'aide afin de l'orienter dans le dialogue. Des invites préliminaires dès le début de l'interaction favorisent la conformité des futures requêtes à un modèle d'énoncé correct[83] tandis que des invites d'aide dynamiques peuvent être réclamées par l'utilisateur ou dispensées automatiquement selon les erreurs rencontrées et leur fréquence.

Modalités alternatives et efficacité

- *Disponibilité de modalités alternatives :* Proposer des modalités supplémentaires tant en entrée qu'en sortie permet de garantir une meilleure accessibilité du système. En cela, le système devrait intégrer des moyens redondants d'obtenir ou de fournir de l'information via des modalités alternatives. La difficulté est d'éviter que l'ajout d'une autre forme de communication n'entraîne une surcompensation des limites de l'interaction vocale au point que les utilisateurs la privilégient[24].
- *Personnalisation de la synthèse vocale :* Il devrait être possible de personnaliser la synthèse vocale selon les préférences de l'utilisateur. De nombreux facteurs peuvent être envisagés afin d'améliorer l'efficacité de l'interaction tels que le rythme de la diction, le volume sonore, le degré de verbosité, et l'intonation. D'autres facteurs plus hédoniques peuvent être également considérés comme le genre de la voix, la présence de formules de politesse, la personnalité de l'agent conversationnel, etc.

Conscience de l'utilisateur

- *Adaptation du système selon les performances de l'utilisateur* : Le système devrait capturer, compiler et traiter les informations relatives aux performances de l'utilisateur afin d'adapter le comportement de l'interface vocale à ses aptitudes[82].
- *Prise en charge de l'expérience passée des utilisateurs avec les agents conversationnels* : La familiarité de l'utilisateur avec d'autres interfaces vocales conditionne ses attentes et contribue à créer un modèle mental unique de l'interaction vocale. La conception d'une interface vocale devrait prendre en compte la manière dont l'expérience avec les principaux assistants personnels intelligents affecte les attentes, les comportements et la satisfaction des utilisateurs[14].
- *Compréhension des besoins des utilisateurs* : Pour mieux comprendre les besoins des utilisateurs, la conception de l'interface vocale devrait inclure dans les différentes étapes du processus les populations ciblées afin de mieux cerner leurs attentes et leurs aptitudes[1].
- *Promotion de l'indépendance sur le système* : Les systèmes basés sur l'interaction vocale peuvent donner l'impression aux utilisateurs d'être assistés et dépendants d'une solution technologique. L'interaction vocale devrait être un outil d'aide aux populations à besoins spécifiques qui évolue en fonction de leurs aptitudes et de la perception qu'ils ont de leur condition. Un niveau d'assistance progressif devrait être prévu pour familiariser les utilisateurs avec le système et surpasser leurs craintes.

Conscience du contexte d'utilisation

- *Adaptation du système selon le contexte d'utilisation* : L'environnement autour de l'utilisateur peut avoir une influence significative sur son expérience de l'interaction vocale et nécessiter une réaction appropriée du système. Dans certains contextes, l'interface vocale devrait être adaptée pour proposer une interaction adéquate et satisfaisante pour l'utilisateur. Plusieurs facteurs situationnels sont à prendre en compte tels que l'interaction dans les milieux publics ou les milieux bruyants.

Exemple : Dans certains contextes, le fait de fournir des modes «silencieux» ou «discret», dans lequel les détails personnels sont soit énoncés de manière plus discrète, soit passés sous silence, peut aider à préserver l'intimité des utilisateurs[1].

Confidentialité et sécurité

- *Minimisation des compromis en matière de sécurité et de confidentialité* : À l'instar du comportement de grande compagnies qui admettent écouter les enregistrements vocaux dans un souci d'amélioration de leurs services¹⁰, certaines fonctionnalités et commodités de l'interaction vocale sont mises en oeuvre au détriment de la protection des données privées de l'utilisateur. Les concepteurs du système devraient trouver de meilleur compromis que ceux de la confidentialité et la sécurité afin de proposer leurs services aux utilisateurs[86].

Exemple : Certains choix techniques en matière de conception peuvent réduire les risques sans impacter de manière significative les fonctionnalités, comme par exemple ne pas être obligé d'être connecté aux services en ligne pour interagir vocalement.

- *Interruption de l'enregistrement des conversations* : L'utilisateur devrait être informé lorsque ses conversations sont enregistrées et avoir la possibilité d'interrompre à tout moment leur enregistrement.

10. Google[76] et Amazon[16] entre autres

Exemple : Le bouton physique sur l'appareil pour couper manuellement le microphone pourrait permettre à l'utilisateur de s'assurer que ses conversations ne sont pas écoutées. Dans le cas contraire, un signal adapté, comme sur les caméras, pourrait informer l'utilisateur que l'enregistrement est en cours.

- *Sensibilisation aux problèmes de sécurité et de confidentialité* : Les utilisateurs ne sont pas assez sensibilisés aux risques pour leurs données personnelles et leur sécurité à l'usage d'une interface vocale. Un support didactique devrait être proposé aux nouveaux utilisateurs pour les conscientiser aux risques existants et aux bonnes pratiques à adopter pour s'en prémunir.

Exemple : Certaines interfaces vocales, telles qu'Alexa, permettent d'effectuer des achats en ligne sans la moindre forme de confirmation de la part de l'utilisateur. Ce faisant, n'importe qui peut passer une commande oralement à la place de l'utilisateur principal. Néanmoins, Un code vocal facultatif peut être paramétré mais est, par défaut, désactivé.

Confiance et fiabilité

- *Prise en charge de commandes longues et complexes* : Le système devrait être mieux adapté aux commandes plus longues et complexes des utilisateurs. Pour certaines populations, la saisie de commandes complexes à l'aide de l'interaction vocale est une alternative plus simple que les saisies à l'aide de méthodes traditionnelles[1].

Exemple : En augmentant la limite de temps pour la formulation de requêtes verbales, le système pourrait améliorer la prise en charge des commandes plus longues.

- *Transparence de l'usage des données personnelles de l'utilisateur* : L'amélioration de la confiance des utilisateurs passe par une meilleure transparence à leur égard. Les utilisateurs devraient être clairement informés de l'usage qui est fait de leurs conversations et données personnelles. Ils devraient également avoir la possibilité d'en gérer l'accès et la conservation[15].
- *Retour de confirmation* : Afin de maintenir la confiance de l'utilisateur et d'améliorer son modèle mental de l'interaction vocale, un retour de confirmation devrait être dispensé par le système afin de lui signifier que sa requête est en bonne voie de traitement[85]. Dans le cas contraire, le système devrait l'informer des raisons pour lesquelles la requête a échoué.

Humanité du système

- *Réalisme contenu de l'humanité du système* : Les utilisateurs sont désireux d'interagir avec un système aux caractéristiques humaines[58, 57, 36, 14]. Néanmoins, ils rejettent une représentation trop réaliste de l'interface vocale provoquant un sentiment de malaise¹¹[57, 11]. La conception de l'interface vocale devrait privilégier certaines caractéristiques humaines sans pour autant être trop réaliste.

Exemple : Afin de rencontrer les souhaits des utilisateurs, le système pourrait entretenir un échange conversationnel à l'aide d'une voix naturelle, à la manière d'un dialogue entre êtres humains.

Intégration du système

- *Meilleure intégration de l'interaction vocale au quotidien* : Les utilisateurs désirent contrôler un nombre plus élevé de dispositifs et d'applications via l'interface vocale[1, 36]. Dans

11. À mettre en corrélation avec "La Vallée de l'étrange"[7]

certains cas, l'interaction vocale peut être la seule forme de technologie permettant aux personnes de rester indépendantes[1]. La centralisation du contrôle des différents dispositifs et applications est également un facteur améliorant leur accessibilité. La conception devrait favoriser l'interopérabilité et une meilleure intégration avec les solutions existantes en matière de domotique mais également avec les applications favorites des utilisateurs.

Exemple : La prise en charge de plusieurs protocoles différents de communication tels que le Wi-Fi ou le Bluetooth permet de contrôler un éventail plus large de dispositifs connectés par la parole.

Chapitre 5

Conclusion

5.1 Limitations

5.1.1 Limitations des études sélectionnées

Malgré leur faible nombre, les études sélectionnées présentent un manque d'homogénéité. Elles diffèrent en matière de configuration et d'environnement de test, de populations ciblées, de méthodologie et d'objectifs.

Configuration et environnement de test : Outre l'absence d'étude longitudinale, aucune expérimentation n'a été conduite in-situ. Au lieu de cela, six études ont été réalisées dans des reproductions d'habitat connecté("living labs")[36, 83, 75, 57, 4] dont l'une était une reproduction en 3D[18]. Deux études concernent l'évaluation d'une application avec une interface vocale[85, 14] et ne nécessitent donc pas d'environnement spécifique. Les autres études sont soit des interviews[58, 1], soit des enquêtes[11, 4], soit des expérimentations avec des données pré-enregistrées[25, 77]. Il est à noter que trois études ont eu recours à un "Wizard-of-Oz"[57, 18, 83] et que seules six études n'utilisent que la parole et la synthèse vocale comme modalités d'interaction[77, 36, 85, 58, 75, 83].

Populations ciblées : Les expérimentations des différentes études n'incluent qu'un faible nombre de participants, entre cinq et seize ($M=8,77$). Certaines populations ne sont pas représentées comme les personnes ayant des difficultés d'audition ou des handicaps lourds. Alors que la majorité des études ciblent les personnes âgées, celles-ci sont encore parfaitement autonomes dans les différentes évaluations.

Apprentissage des commandes vocales : Aucune étude ne s'est focalisée sur les utilisateurs interagissant pour la première fois avec un système à l'aide de la voix. Cela est sans doute à mettre en rapport avec la quasi-absence de préoccupations relatives à la découverte des commandes vocales. Le sujet n'est traité que dans une seule étude[14] sans pour autant que les participants objectent des difficultés à cet égard : après un bref tutoriel, tous les participants sont parvenus à formuler des requêtes et utiliser le système. Seul un article mentionne des problèmes d'apprentissage des commandes[58].

Le manque de considération pour cet aspect de l'interaction vocale laisse supposer qu'il existe des biais de sélection dans le choix des participants aux différentes études. Dans cinq études[58, 1, 83, 36, 4], il est avéré que les participants ont des affinités avec la technologie vocale, voire connaissent l'interface utilisée pendant l'expérimentation ou toute autre interface similaire. Dans l'une des études, les proches des participants ont été invités à participer aux évaluations[57] sans que les chercheurs ne puissent exclure leur influence sur les résultats obtenus. Les autres études sont plus permissives quant aux requêtes formulées[75, 18] ou au contraire bien trop dirigiste que pour laisser une marge de manoeuvre à l'utilisateur[85]. Les études non-mentionnées[77, 11, 25] ne contiennent pas d'expérimentation avec des participants.

5.1.2 Limitations des recommandations de conception

Validation des recommandations : La finalité de ce mémoire est de proposer un ensemble de recommandations de conception propres à l'interaction vocale pour les populations à besoins spécifiques. Afin de valider l'utilité et l'exhaustivité des recommandations, un prototype ainsi qu'une expérimentation avec plusieurs populations différentes seraient nécessaires.

Des recommandations abstraites : Les recommandations proposées ne répondent pas à des problèmes concrets avec des solutions directement applicables. Ce sont avant tout des lignes directrices pouvant déboucher sur diverses solutions à devoir mettre en pratique au cas par cas.

Temporalité de la technologie : L'interaction vocale connaît un formidable essor cette dernière décennie. La technologie évolue vite ainsi que les habitudes, les attentes et les besoins des utilisateurs. Il n'est pas possible de prévoir à l'avance si les recommandations seront toujours d'actualité dans un futur relativement proche. De même, ces recommandations ont été rédigées sur base d'études qui sont également datées et peuvent ne plus représenter les préoccupations actuelles des utilisateurs. Par exemple, l'une des études[85] est basée sur un système de réservation par téléphone avec répondeur automatique. Une pratique qui tend à disparaître et avec elle certains comportements.

5.2 Remarques finales et perspectives futures

L'interaction présente de nombreux avantages dans les domaines de la domotique[32, 56] et des technologies d'assistance en complément d'autres modalités[53, 30]. Néanmoins il n'existe pas de cadre formel ayant pour but de guider les concepteurs dans la réalisation d'une interface vocale garantissant une expérience utilisateur satisfaisante pour des populations à besoins spécifiques. Pour répondre à ce manque, ce mémoire a permis de créer un ensemble d'heuristiques applicables au domaine de l'interaction vocale à l'aide des travaux de Rusu *et al.*[63], Nielsen *et al.*[50] et Arhippainen[2], d'identifier les préoccupations des populations à besoins spécifiques en matière d'interaction vocale, et de proposer des recommandations pour chaque heuristique en se basant sur les principes de Wobbrock *et al.*[81]. Les recommandations de conception répondent à l'ensemble des besoins mentionnés dans les études tout en étant limitées par celles-ci. Ce sont des lignes directrices qui ne contraignent pas les concepteurs dans une solution spécifique. Elles se veulent simples et orientées vers une meilleure accessibilité et adaptabilité aux besoins des utilisateurs.

Afin d'aller plus loin dans la démarche, plusieurs pistes sont à explorer. La mise en évidence des préoccupations des utilisateurs a montré une absence de considération pour les problématiques liées à l'apprentissage et la découverte des commandes vocales. Étudier une telle question, notamment avec des participants qui n'ont pas d'affinités ni de connaissances en matière de technologie vocale, pourrait conduire à une meilleure compréhension des mécanismes d'apprentissage et de découverte des commandes vocales par des populations à besoins spécifiques. Au travers de la lecture des différents articles, il est également apparu que les participants étaient moins sensibles au respect de leur intimité et de la confidentialité de leurs données. Pourtant, la technologie vocale se veut par nature intrusive et indiscreète. Cette question pourrait être abordée plus en profondeur tant elle semble d'actualité¹². Enfin, et surtout, une validation des recommandations par une seconde étude pourrait renforcer leur légitimité voire assurer leur complétude.

12. Après les révélations des écoutes des conversations privées, Apple et Amazon font machine arrière[34].

Références

- [1] Ali Abdolrahmani, Ravi Kuber, and Stacy M. Branham. ““Siri Talks at You”: An Empirical Investigation of VoiceActivated Personal Assistant (VAPA) Usage by Individuals Who Are Blind”. In: *ASSETS '18 Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 2018, pp. 249–258.
- [2] Leena Arhippainen. “A Tutorial of Ten User Experience Heuristics”. In: *AcademicMindTrek '13 Proceedings of International Conference on Making Sense of Converging Media*. 2013.
- [3] Matthew P. Aylett et al. “None of a CHInd: Relationship Counselling for HCI and Speech Technology”. In: *CHI EA '14 CHI '14 Extended Abstracts on Human Factors in Computing Systems*. 2014, pp. 749–760.
- [4] Shiri Azenkot and Nicole B. Lee. “Exploring the Use of Speech Input by Blind People on Mobile Devices”. In: *ASSETS '13 Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. 2013.
- [5] Fabio Ballati, Fulvio Corno, and Luigi De Russis. ““Hey Siri, do you understand me?": Virtual Assistants and Dysarthria”. In: *WoRIE '18 Proceedings of the 7th International Workshop on the Reliability of Intelligent Environments*. 2018.
- [6] Scott Beach et al. “Older Adults and Voice Interaction: A Pilot Study with Google Home”. In: *ACM Transactions on Accessible Computing (TACCESS)* (2009).
- [7] Lydia Ben Ytzhak. *Petit détour par la vallée de l'étrange*. URL: <https://lejournal.cnrs.fr/articles/petit-detour-par-la-vallee-de-letrange>.
- [8] A.J. Bernheim Brush et al. “Home Automation in the Wild: Challenges and Opportunities”. In: *CHI '11 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2011, pp. 2115–2124.
- [9] Nigel Bevan et al. “New ISO Standards for Usability, Usability Reports and Usability Measures”. In: *Proceedings of the 18th International Conference on Human-Computer Interaction. Theory, Design, Development and Practice*. 2016, pp. 268–278.
- [10] The Amazon Blog. *How Alexa keeps getting smarter*. URL: <https://blog.aboutamazon.com/devices/how-alexa-keeps-getting-smarter>.
- [11] Zoraida Callejas and Ramón López-Cózar. “Designing smart home interfaces for the elderly”. In: *ACM SIGACCESS Accessibility and Computing archive* (2009), pp. 10–16.
- [12] Clare Carroll et al. “Robin: Enabling Independence For Individuals With Cognitive Disabilities Using Voice Assistive Technology”. In: *CHI EA '17 Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2017, pp. 46–53.
- [13] Philip R. Cohen and Sharon L. Oviatt. “The Role of Voice in Human-machine Communication”. In: *Voice Communication Between Humans and Machines*. 1994, pp. 34–75.

- [14] Eric Corbett and Astrid Weber. “What can I say ? Addressing user experience challenges of a mobile voice user interface for accessibility”. In: *MobileHCI '16 Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 2016, pp. 72–82.
- [15] Benjamin Cowan et al. “What Can I Help You With ? : Infrequent Users Experiences of Intelligent Personal Assistants”. In: *MobileHCI '17 Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 2017.
- [16] Matt Day, Giles Turner, and Natalia Drozdak. *Amazon Workers Are Listening to What You Tell Alexa*. URL: <https://www.bloomberg.com/news/articles/2019-04-10/is-anyone-listening-to-you-on-alexa-a-global-team-reviews-audio>.
- [17] George Demiris et al. “Older adults’ attitudes towards and perceptions of “smart home” technologies: a pilot study”. In: *Medical Informatics and the Internet in Medicine* (2004), pp. 87–84.
- [18] Jan Derboven, Jonathan Huyghe, and Dirk De Grooff. “Designing Voice Interaction for People with Physical and Speech Impairments”. In: *NordiCHI '14 Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. 2014, pp. 217–226.
- [19] Department of Economic and Social Affairs. *World Population Prospects (The 2017 Revision)*. Tech. rep. United Nations, 2017. URL: https://esa.un.org/unpd/wpp/Publications/Files/WPP2017_KeyFindings.pdf.
- [20] United Nation - Departement of Economic and Social Affairs. *Convention on the Rights of Persons with Disabilities (CRPD)*. URL: <https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html>.
- [21] Eurostat. *Personnes déclarant une difficulté dans les activités domestiques par sexe, âge et nécessité d’assistance*. 2012. URL: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=hlth_dpeh120&lang=fr.
- [22] Eurostat. *Population par sexe, âge et statut d’incapacité*. 2012. URL: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=hlth_dpeh005&lang=fr.
- [23] Jinjuan Feng et al. “Speech-based Navigation and Error Correction: A Comprehensive Comparison of Two Solutions”. In: *Universal Access in the Information Society* (2011), pp. 17–31.
- [24] Anushay Furqan, Chelsea Myers, and Jichen Zhu. “Learnability Through Adaptive Discovery Tools in Voice User Interfaces”. In: *CHI EA '17 Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2017, pp. 1617–1623.
- [25] Jort Gemmeke et al. “Self-taught assistive vocal interfaces: An overview of the ALADIN project”. In: *INTERSPEECH '13 Annual Conference of the International Speech Communication Association*. 2013, pp. 2038–2043.
- [26] Debjyoti Ghosh et al. “Proceedings of the Sixth International Symposium of Chinese CHI (ChineseCHI '18)”. In: *Mobile HCI '18 Proceedings of the 20th Conference on Human-Computer Interaction with Mobile Devices*. 2018, pp. 11–15.
- [27] Peter Gregor, Alan F. Newell, and Mary Zajicek. “Designing for dynamic diversity: interfaces for older people”. In: *Assets '02 Proceedings of the 5th International ACM Conference on Assistive Technologies*. 2002, pp. 151–156.
- [28] Nielsen Norman Group. *10 Usability Heuristics for User Interface Design*. URL: <https://www.nngroup.com/articles/ten-usability-heuristics/>.

- [29] Nielsen Norman Group. *The Definition of User Experience (UX)*. URL: <https://www.nngroup.com/articles/definition-user-experience/>.
- [30] Foad Hamidi et al. "CanSpeak: A Customizable Speech Interface for People with Dysarthric Speech". In: *ICCHP '10: Computers Helping People with Special Needs*. 2010, pp. 605–612.
- [31] Mark Hawley et al. "A speech-controlled environmental control system for people with severe dysarthria". In: *Medical engineering physics* (2007).
- [32] Jonathan Huyghe, Jan Derboven, and Dirk De Grooff. "ALADIN: Adaptive Voice Interface for People with Disabilities". In: *NordiCHI '14 Proceedings of the 8th Nordic Conference on Human-Computer Interaction*. 2014.
- [33] Apple Inc. *Page d'accueil de Siri*. URL: <https://www.apple.com/fr/siri/>.
- [34] The Independent. *Apple And Google Stop Workers From Listening To Intimate Recordings Of Customers*. URL: <https://www.independent.co.uk/life-style/gadgets-and-tech/news/apple-siri-private-recordings-google-customers-users-a9036681.html>.
- [35] Min-Soo Kang, Kyung Mi Kim, and Hee-Cheol Kim. "A Questionnaire Study for the Design of Smart Home for the Elderly". In: *HEALTHCOM '06 Proceedings of the 8th International Conference on e-Health Networking, Applications and Services*. 2006.
- [36] Jarosław Kowalski et al. "Older Adults and Voice Interaction: A Pilot Study with Google Home". In: *CHI EA '19 Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019.
- [37] Richard Ladner. "Design for User Empowerment". In: *Interactions* (2015).
- [38] Lorna Lines and Kate S. Hone. "Multiple voices, multiple choices : Older adults' evaluation of speech output to support independent living". In: *Gerontechnology* (2006), pp. 78–91.
- [39] Dima Litvak, Yaniv Zigel, and Israel Gannot. "Fall detection of elderly through floor vibrations and sound". In: *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2008.
- [40] Ramón López-Cózar and Zoraida Callejas. "Multimodal Dialogue for Ambient Intelligence and Smart Environments". In: *Handbook of Ambient Intelligence and Smart Environments*. Springer US, 2010, pp. 559–579.
- [41] Ewa Luger and Abigail Sellen. "'Like Having a Really bad PA': The Gulf between User Expectation and Experience of Conversational Agents". In: *CHI '16 Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2016, pp. 5286–5297.
- [42] Michal Luria, Guy Hoffman, and Oren Zuckerman. "Comparing Social Robot, Screen and Voice Interfaces for Smart-Home Control". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2017, pp. 580–628.
- [43] Frans Mäyrä et al. "Probing a Proactive Home : Challenges in Researching and Designing Everyday Smart Environments". In: *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments* (2006).
- [44] Alessandra Mileao, Roberto Bisiani, and Davide Merico. "Reasoning Support for Risk Prediction and Prevention in Independent Living". In: *Proceedings of The 24th International Conference on Logic Programming*. 2009, pp. 361–395.
- [45] Sebastian Möller, Florian Gödde, and Maria Wolters. "Corpus Analysis of Spoken Smart-Home Interactions with Older Users". In: *LREC '08 Proceedings of the International Conference on Language Resources and Evaluation*. 2008.

- [46] Christine Murad et al. “Design guidelines for hands-free speech interaction”. In: *Mobile HCI '18 Proceedings of the 20th Conference on Human-Computer Interaction with Mobile Devices*. 2018.
- [47] Chelsea Myers, Anushay Furqan, and Jichen Zhu. “The Impact of User Characteristics and Preferences on Performance with an Unfamiliar Voice User Interface”. In: *CHI '19 Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019.
- [48] Chelsea Myers et al. “Patterns for How Users Overcome Obstacles in Voice User Interfaces”. In: *CHI '18 Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018.
- [49] Alan F. Newell. “Extra-ordinary Human-computer Interaction”. In: *Extra-ordinary Human-computer Interaction : Interfaces for Users with Disabilities*. 1995, pp. 3–18.
- [50] Jakob Nielsen and Rolf Molich. “Heuristic Evaluation of User Interfaces”. In: *CHI '90 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1990, pp. 249–256.
- [51] Organisation internationale de normalisation. *Personnes déclarant une difficulté dans les activités domestiques par sexe, âge et nécessité d'assistance*. URL: <https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-1:v1:fr>.
- [52] Cathy Pearl. *Designing Voice User Interfaces: Principles of Conversational Experiences*. O'Reilly Media, Inc., 2016.
- [53] Carlos Galinho Pires et al. “On the Benefits of Speech and Touch Interaction with Communication Services for Mobility Impaired Users”. In: *AAL '11 Proceedings of the 1st International Living Usability Lab Workshop on AAL Latest Solutions, Trends and Applications*. 2011.
- [54] Martha E. Pollack. “Intelligent Technology for an Aging Population : The Use of AI to Assist Elders with Cognitive Impairment”. In: *AI Magazine* (2005), pp. 9–24.
- [55] Martin Porcheron, Joel E. Fischer, and Sarah Sharples. “"Do Animals Have Accents?": Talking with Agents in Multi-Party Conversation”. In: *CSCW '17 Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 2017, pp. 207–219.
- [56] Martin Porcheron et al. “Voice Interfaces in Everyday Life”. In: *CHI '18 Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018.
- [57] François Portet et al. “Design and evaluation of a smart home voice interface for the elderly : acceptability and objection aspects”. In: *Personal and Ubiquitous Computing* (2013), pp. 127–144.
- [58] Alisha Pradhan, Kanika Mehta, and Leah Findlater. “« Accessibility came by accident » : use of voice-controlled intelligent personal assistants by people with disabilities”. In: *CHI '18 Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018.
- [59] Amanda Purington et al. ““Alexa is my new BFF”: Social Roles, User Satisfaction, and Personification of the Amazon Echo”. In: *CHI EA '17 Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2017, pp. 2853–2859.
- [60] Patrick Rabbitt and Mike Anderson. “The Lacunae of Loss? Aging and the Differentiation of Cognitive Abilities”. In: *Lifespan Cognition: Mechanisms of Change*. Oxford University Press, 2006, pp. 331–343.

- [61] Vincent Rialle et al. “What do family caregivers of Alzheimer’s disease patients desire in smart home technologies? Contrasted results of a wide survey”. In: *Methods of Information in Medicine* (2008), pp. 63–69.
- [62] Frank Rudzicz et al. “Speech Interaction with Personal Assistive Robots Supporting Aging at Home for Individuals with Alzheimer’s Disease”. In: *ACM Transactions on Accessible Computing (TACCESS) - Special Issue on Speech and Language Processing for AT (Part 3)* (2015).
- [63] Cristian Rusu et al. “A Methodology to Establish Usability Heuristics”. In: *ACHI ’11 Proceedings of the 4th International Conference on Advances in Computer-Human Interactions*. 2011.
- [64] Privender Saini et al. “Benefits of Social Intelligence in Home Dialogue Systems”. In: *INTERACT’05 Proceedings of the 2005 IFIP TC13 international conference on Human-Computer Interaction*. 2005, pp. 510–521.
- [65] Andrew Sears and Mark Young. “Physical Disabilities and Computing Technologies: An Analysis of Impairments”. In: *The Human-computer Interaction Handbook*. 2003, pp. 482–503.
- [66] Ben Shneiderman. “The limits of speech recognition”. In: *Communications of the ACM* (2000).
- [67] Pavel Slavík, Vladislav Němec, and Adam J. Sporka. “Speech Based User Interface for Users with Special Needs”. In: *TSD ’05 Proceedings of the 8th International Conference on Text, Speech and Dialogue*. 2005, pp. 45–55.
- [68] Christoph Stahl and Pascal Laub. “Maintaining multiple sclerosis patients’ quality of life – a case study on environment control assistance in a smart home”. In: *PETRA ’17 Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments*. 2017, pp. 83–86.
- [69] Lisa Stifelman, Adam Elman, and Anne Sullivan. “Designing Natural Speech Interactions for the Living Room”. In: *CHI EA ’13 CHI ’13 Extended Abstracts on Human Factors in Computing Systems*. 2013, pp. 1215–1220.
- [70] António Teixeira et al. “Speech as the Basic Interface for Assistive Technology”. 2009.
- [71] António Teixeira et al. “Speech-centric Multimodal Interaction for Easy-to-access Online Services – A Personal Life Assistant for the Elderly”. In: *Proceedings of the 5th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*. 2013.
- [72] Bui Huu Trung. *Multimodal Dialogue Management - State of the art*. Centre for Telematics and Information Technology (CTIT), 2006.
- [73] Nations Unies. *Faits et chiffres sur le handicap*. URL: <https://www.un.org/development/desa/disabilities-fr/faits-et-chiffres-sur-le-handicap.html>.
- [74] Michel Vacher et al. “Experimental Evaluation of Speech Recognition Technologies for Voice-based Home Automation Control in a Smart Home”. In: *Proceedings of the 4th Workshop on Speech and Language Processing for Assistive Technologies*. 2013, pp. 99–105.
- [75] Michel Vacher et al. “Evaluation of a Context-Aware Voice Interface for Ambient Assisted Living: Qualitative User Study vs. Quantitative System Evaluation”. In: *ACM Transactions on Accessible Computing (TACCESS) - Special Issue on Speech and Language Processing for AT* (2015).

- [76] Lente Van Hee et al. *Google employees are eavesdropping, even in your living room, VRT NWS has discovered*. URL: https://www.vrt.be/vrtnws/en/2019/07/10/google-employees-are-eavesdropping-even-in-flemish-living-rooms/?mod=article_inline.
- [77] Ravichander Vippera et al. "Speech Input from Older Users in Smart Environments: Challenges and Perspectives". In: *UAHCI '09 Proceedings of the 5th International on Conference Universal Access in Human-Computer Interaction. Part II: Intelligent and Ubiquitous Interaction Environments*. 2009, pp. 117–126.
- [78] John A. Waterworth et al. "Ageing in a networked society: social inclusion and mental stimulation". In: *PETRA '09 Proceedings of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments*. 2009, pp. 1–5.
- [79] Mark Weiser. "The computer for the 21st century". In: *Scientific American* (1991), pp. 94–104.
- [80] Katherine Wild et al. "Unobtrusive In-Home Monitoring of Cognitive and Physical Health: Reactions and Perceptions of Older Adults". In: *Journal of Applied Gerontology* (2008), pp. 181–200.
- [81] Jacob O. Wobbrock et al. "Ability-Based Design: Concept, Principles and Examples". In: *ACM Transactions on Accessible Computing (TACCESS)* (2011).
- [82] Jacob O. Wobbrock et al. "Ability-based Design". In: *Communications of the ACM* (2018), pp. 62–71.
- [83] K. Maria Wolters et al. "Making It Easier for Older People to Talk to Smart Homes: The Effect of Early Help Prompts". In: *Universal Access in the Information Society* (2010), pp. 311–325.
- [84] Nicole Yankelovich. "How do users know what to say?" In: *Interactions* (1996).
- [85] Mary Zajicek, Richard Wales, and Andrew Lee. "Speech interaction for older adults". In: *Universal Access in the Information Society* (2004), pp. 122–130.
- [86] Eric Zeng, Shrirang Mare, and Franziska Roesner. "End User Security and Privacy Concerns with Smart Homes". In: *SOUPS '17 Proceedings of the Thirteenth USENIX Conference on Usable Privacy and Security*. 2017, pp. 65–80.
- [87] Martina Ziefle and wiktoria Wilkowska. "Technology acceptability for medical assistance". In: *Proceedings of the 4th International Conference on Pervasive Computing Technologies for Healthcare*. 2010, pp. 1–9.

Annexe A

Annexe : Préoccupations des populations à besoins spécifiques

Étude	Heuristique	Groupe ciblé	Évaluation/Constataion
[83]	VOCH1	Population âgée	Les utilisateurs âgés ayant un style d'interaction plus adéquat (ratio de mots-clés et vocabulaire partagé plus élevés) ont tendance à avoir une impression générale du système plus favorable.
[83]	VOCH6 + VOCH7	Population âgée	Les utilisateurs plus âgés qui ont d'abord été exposés à l'aide au préalable sont plus susceptibles de choisir un style d'interaction efficace que les utilisateurs ayant reçu une aide dynamique. Cet effet n'est pas vérifié sur les jeunes utilisateurs qui choisissent toujours un style d'interaction adéquat.
[58]	VOCH10	Population malvoyante	Les participants (6/16) craignent que le système ne soit toujours à l'écoute et enregistre leurs propos.
[58]	VOCH10	Population malvoyante	Les participants (6/16) ont exprimé des préoccupations en matière de confidentialité concernant la collecte de données personnelles.
[58]	VOCH10	Population malvoyante	Les participants (4/16) ont soulevé des problèmes de sécurité tels que ceux liés au piratage des services hébergés dans le cloud.
[58]	VOCH2 + VOCH8	Population malvoyante	Les participants (2/16) ont mentionné un manque d'accessibilité visuelle lié à la conception physique du périphérique (par exemple, des signaux lumineux non perceptibles par les personnes malvoyantes ou des boutons physiques avec des inscriptions peu contrastées).
[58]	VOCH7	Population malvoyante	La modalité d'entrée de préférence des utilisateurs est la parole, mais une majorité d'entre eux (9/16) désirent aussi d'autres moyens d'interaction tels qu'une télécommande, un smartphone, une smartwatch, la détection gestuelle ou le contact direct pour contrôler l'appareil.
[58]	VOCH13 + VOCH7	Population malvoyante	Selon certains participants (10/16), les assistants personnels intelligents ont remplacé de nombreux appareils et applications disparates. Ils ont amélioré l'efficacité et le sentiment d'indépendance des utilisateurs pour toute une variété de tâches.
[58]	VOCH1	Population malvoyante	Les participants (10/16) ont indiqué que la configuration du périphérique était difficile, à la fois lors de l'achat initial ou lorsque l'appareil est déconnecté du réseau.
[58]	VOCH13	Population malvoyante	Les participants (8/16) ont mentionné des problèmes d'appairage avec l'application mobile.
[58]	VOCH5	Population malvoyante	Les participants (8/16) ont déclaré éprouver des difficultés à découvrir les fonctionnalités existantes ou ont mentionné qu'ils désiraient une fonctionnalité particulière pourtant existante - démontrant de la sorte le problème.

Tableau A.1 – Préoccupations des populations à besoins spécifiques en matière d'interaction vocale

Étude	Heuristique	Groupe ciblé	Évaluation/Constatation
[57]	VOCH1 + VOCH7	Population âgée	Bien que la plupart des personnes âgées (5/8) aient spontanément contrôlé l'environnement de test en prononçant des phrases, par la suite, la majorité (5/8) ont déclaré vouloir contrôler leur habitat en utilisant des mots clés. Ils pensent que ce mode d'interaction serait plus rapide et efficace.
[57]	VOCH12	Population âgée	Seulement la moitié des personnes âgées (4/8) ont trouvé naturel d'interagir à l'aide de la parole même si certains ont admis parler à voix haute quand ils sont seuls.
[57]	VOCH12	Population âgée	La plupart des personnes âgées (6/8) ont déclaré préférer s'adresser à l'environnement plutôt qu'à une télécommande ou à un compagnon robotique. Il faut cependant noter qu'ils n'ont jamais interagi avec une entité robotisée.
[57]	VOCH4	Population âgée	En cas de malentendu, la plupart des personnes âgées (6/8) préfèrent que le système leur demande de répéter plutôt que de proposer un choix entre plusieurs solutions.
[57]	VOCH8	Population âgée	Presque toutes les personnes âgées (6/8) n'apprécient pas d'être informées par le système d'un rendez-vous imminent car cela leur donne le sentiment d'être assistées. Elles craignent de perdre certaines de leurs capacités intellectuelles par manque de pratique.
[57]	VOCH8	Population âgée	Toutes les personnes âgées, à l'exception d'une (7/8), ont apprécié la possibilité de contrôler l'environnement à l'aide de commandes vocales. cependant, ces mêmes personnes âgées ne prévoient pas d'utiliser ce système pour le moment. Elles pensent que cette façon d'interagir est dédiée aux personnes handicapées. Les personnes âgées ont peur d'être considérées comme des personnes dépendantes si elles décidaient d'adopter ce système chez elles. Elles craignent également de perdre leur autonomie à cause du système.
[57]	VOCH10	Population âgée	En ce qui concerne la sécurité, toutes les personnes âgées ont apprécié que le système signale des problèmes de sécurité. Elles admettent qu'elles ont tendance à oublier de vérifier si la porte est fermée ou le gaz est coupé.
[57]	VOCH10	Population âgée	La plupart des participants ont estimé que le système évalué ne mettait pas leur vie privée en danger. En fait, peu d'âinés craignent d'être enregistrés. En revanche, presque toutes les personnes âgées (7/8) ne veulent pas de caméra dans leur domicile.
[57]	VOCH2	Population âgée	Presque toutes les personnes âgées (7/8) préfèrent que le système émette un son bref avant de se manifester de manière inopinée. Cela leur permettrait d'être plus attentifs au message diffusé et de distinguer celui-ci du son de la télévision ou d'un proche faisant irruption dans la pièce. Pour cinq participants (5/8), le son doit être différent en fonction de l'alerte.

Tableau A.2 – Préoccupations des populations à besoins spécifiques en matière d'interaction vocale

Étude	Heuristique	Groupe ciblé	Évaluation/Constatation
[57]	VOCH12	Population âgée	Les participants ont unanimement exprimé le fait que la voix du système devait être naturelle et non synthétique.
[14]	VOCH1	Population avec trouble de la dextérité manuelle	Les participants s'attendent à une écoute continue du système même lorsqu'ils font des pauses dans l'interaction pour réfléchir.
[14]	VOCH1	Population avec trouble de la dextérité manuelle	Quelle que soit la forme d'interaction la plus appropriée au contexte d'utilisation (style conversationnel ou injonctif), le style conversationnel est préféré et souhaité du moins du point de vue de l'expérience utilisateur.
[14]	VOCH8	Population avec trouble de la dextérité manuelle	Les utilisateurs assument le modèle mental d'interaction à partir d'autres agents conversationnels populaires.
[14]	VOCH5	Population avec trouble de la dextérité manuelle	Tous les utilisateurs sont parvenus à naviguer dans l'interface graphique à l'aide de la parole après avoir complété l'introduction.
[14]	VOCH5	Population avec trouble de la dextérité manuelle	Les messages didactiques ont raccourci l'expérience d'apprentissage en la déplaçant au sein de l'application et en fournissant des exemples applicables à un usage quotidien.
[75]	VOCH3	Population âgée et population malvoyante	Le ratio de commandes vocales répétées est plus élevé chez les personnes âgées que chez les personnes avec une déficience visuelle (93,6% contre 55,4%). Une corrélation existe entre le taux d'erreur et l'âge des participants : plus les personnes sont âgées, plus le taux d'erreur est élevé.
[75]	VOCH1	Population âgée et population malvoyante	Les participants ayant une déficience visuelle ont strictement suivi la séquence de commandes à exécuter. Ils ont attendu l'exécution de la commande précédente avant d'en énoncer une nouvelle. De plus, leur attention était totalement focalisée sur la tâche en cours et n'ont jamais entrepris d'autre activité simultanément. En revanche, les personnes âgées ont énoncé de nouvelles commandes même si l'action résultant de la précédente commande vocale n'était pas terminée.
[75]	VOCH8	Population âgée et population malvoyante	L'utilisation d'un système de contrôle d'une maison connectée implique deux inconvénients principaux. Premièrement, les aînés ont peur de finir par ne rien faire dans leur maison. Deuxièmement, plusieurs personnes âgées ont déclaré qu'ils pourraient avoir le sentiment de ne pas être capables d'agir seuls. En complément de la peur de devenir dépendants du système, les participants malvoyants ont peur de perdre le contrôle de l'habitat. En raison de leurs déficiences visuelles, ils veulent garder le contrôle et donc ils ont besoin de maîtriser leur habitat pour se sentir en sécurité.

Tableau A.3 – Préoccupations des populations à besoins spécifiques en matière d'interaction vocale

Étude	Heuristique	Groupe ciblé	Évaluation/Constatation
[75]	VOCH12	Population âgée et population malvoyante	Les participants sont restés seuls dans l'environnement de test durant tous les scénarios. Il a été observé que cette configuration (parler dans le vide) était gênante pour les personnes âgées. Ils cherchaient un interlocuteur lors de l'interaction par la parole. À l'inverse, les participants malvoyants n'ont pas cherché d'interlocuteur et n'ont prêté attention qu'à l'exécution de leurs commandes.
[75]	VOCH12	Population âgée et population malvoyante	Les participants ont ressenti la «présence» du système à différents degrés. Les personnes âgées ont le sentiment de ne plus être seules en utilisant ce système, alors que les participants avec des déficiences visuelles avaient l'impression d'être seuls et de parler à voix haute dans le vide.
[75]	VOCH7 + VOCH1	Population âgée et population malvoyante	L'une des principales causes du non-respect de la grammaire était l'absence du mot clé «Nestor». Les personnes âgées se sont souvent adressés au système comme si elles avaient une discussion avec et n'ont donc pas ressenti le besoin de spécifier un nom au début de chaque phrase. Les ajustements effectués par les personnes âgées sont les suivants : ajout de mots de courtoisie, déplacement du mot d'éveil «Nestor» à la fin de la phrase, et utilisation de la forme interrogative pour demander et non ordonner l'exécution d'une tâche.
[75]	VOCH10	Population âgée et population malvoyante	Aucun participant n'a considéré les microphones comme un inconvénient à l'utilisation du système. Deux participants ayant des déficiences visuelles ont indiqué que l'enregistrement des conversations ne devrait pas être accessible hors de l'habitat et qu'il devrait être possible de l'arrêter à tout moment.
[75]	VOCH2	Population âgée et population malvoyante	Un problème commun aux deux groupes est le fait que le système n'a jamais indiqué s'il comprenait les commandes énoncées ou si elles étaient terminées. Cela est dû au manque de retour du système autre que le résultat des actions effectuées.
[18]	VOCH9	Population avec troubles physiques et population avec troubles de la parole	Les participants n'ont pas nommé spécifiquement les objets de l'environnement. Dans une conversation, cette information est déduite du contexte, ou est définie ultérieurement par un échange. Sans forme d'intelligence du système basée sur la localisation ou de système conversationnel, les utilisateurs doivent être spécifiques à propos de leurs intentions.
[18]	VOCH9	Population avec troubles physiques et population avec troubles de la parole	Une variation considérable dans le choix des mots utilisés par les participants pour nommer les objets de l'environnement a été constatée. De plus, le choix des mots est influencé par le point de vue de l'utilisateur. La même porte est appelée la «porte du salon» lorsque l'utilisateur est dans la chambre à coucher, et comme «porte de la chambre» quand l'utilisateur est dans le salon. Ce type d'ambiguïté est encore plus difficile pour un système sans intelligence basée sur la localisation.

Tableau A.4 – Préoccupations des populations à besoins spécifiques en matière d'interaction vocale

Étude	Heuristique	Groupe ciblé	Évaluation/Constatation
[18]	VOCH9	Population avec troubles physiques et population avec troubles de la parole	Le problème le plus courant est le manque de spécificité des commandes verbales conduisant à un message d'erreur. En général, les utilisateurs ont complété la commandes en formulant une nouvelle commande après avoir obtenu de l'aide. Quand ils corrigent la commande immédiatement après le message d'erreur, les participants sont plus susceptibles d'ajouter simplement l'information manquante, sans utiliser de mot d'éveil, et en omettant le restant de la commande. Ce comportement est lié au fait que les utilisateurs s'attendent à ce que le système ait une connaissance de l'ensemble de la conversation.
[18]	VOCH7	Population avec troubles physiques et population avec troubles de la parole	la principale limitation identifiée par les participants avant l'évaluation est le fait que le système doit être formé pour reconnaître les synonymes et formulations alternatives pour qu'ils soient utilisables. Cela diminue potentiellement la précision du système, en particulier dans les premiers temps d'utilisation.
[18]	VOCH1	Population avec troubles physiques et population avec troubles de la parole	Les différentes itérations ont montré que lorsque le système est adressé à l'aide d'un nom, les utilisateurs ont tendance à utiliser un style verbal injonctif assez court.
[18]	VOCH1	Population avec troubles physiques et population avec troubles de la parole	Au quotidien, les commandes vocales s'entremêlent avec les conversations naturelles. Il est donc difficile d'éviter que certains mots ou structures de phrases puissent être interprétés comme des commandes vocales. La plupart des participants ont convenu qu'il était nécessaire d'adresser le système à l'aide d'un nom ou d'un mot clé pour réduire cette ambiguïté. Pour plus de la moitié des participants (5/9), cette ambiguïté serait une raison suffisante pour éteindre le système lorsqu'ils ont de la visite.
[1]	VOCH7	Population malvoyante	Interrogés sur les cas où l'interaction vocale pourrait être utilisée, les participants ont suggéré des scénarios dans lesquels une une réponse rapide était nécessaire. Si l'utilisateur est engagé dans une autre tâche, l'interaction vocale a été pensée pour promouvoir la capacité à effectuer plusieurs tâches.
[1]	VOCH2	Population malvoyante	Les participants ont mentionné que les assistants vocaux devraient permettre aux utilisateurs de personnaliser la synthèse vocale (par exemple, le débit, la clarté et l'intensité de la parole) en fonction de leurs besoins et de la nature de la tâche à accomplir.
[1]	VOCH9 + VOCH11	Population malvoyante	Interrogés sur leurs expériences avec les assistants vocaux, dix participants (10/14) ont souligné que les assistants vocaux mobiles tels que Siri interprètent mal leurs commandes ou requêtes, en particulier dans les lieux publics bruyants. Les participants sont frustrés lorsqu'une action inattendue survient en raison d'une mauvaise interprétation de la commande d'origine. Cela conduit à un manque de confiance dans l'interaction vocale pour effectuer des tâches.

Tableau A.5 – Préoccupations des populations à besoins spécifiques en matière d'interaction vocale

Étude	Heuristique	Groupe ciblé	Évaluation/Constatation
[1]	VOCH11	Population malvoyante	Le manque de confiance dans l'interaction vocale pour s'acquitter des tâches de manière appropriée conduit les participants à développer des stratégies pour vérifier la bonne exécution de leurs commandes.
[1]	VOCH1	Population malvoyante	Pour énoncer une commande vocale, les utilisateurs doivent activer l'assistant vocal soit en appuyant sur un bouton de l'appareil ou à l'aide d'un signal verbal. Lorsque le système est éveillé, une période de temps spécifique est définie pour formuler une commande. Quatre participants (4/14) ont trouvé ce délai d'attente frustrant, en particulier lors de la saisie d'une commande complexe, telle que l'ajout d'un rendez-vous à l'agenda ou la rédaction d'un courriel. Les participants ont mentionné un manque de retour du système quant au temps restant avant la fin de l'intervalle de temps.
[1]	VOCH1 + VOCH9	Population malvoyante	Huit participants (8/14) ont signalé des cas où le système ne répond pas lorsque les mots d'éveil sont prononcés, ou à l'inverse, est activé de manière inattendue. Pour quatre participants (4/14) ces expériences les ont amenés à privilégier l'activation à l'aide d'un bouton pour garder le contrôle sur l'éveil du système. Les participants ont constaté que ce cas se présente plus fréquemment à domicile lorsque la télévision est allumée. Trois participants (3/14) préfèrent éteindre leur assistant vocal dans certaines situations telles que les rassemblements sociaux du fait de la sensibilité du microphone.
[1]	VOCH11	Population malvoyante	Cinq participants (5/14) ont fait part de leur frustration en énonçant des requêtes plus complexes avec certains assistants personnels mobiles car ils ne font que proposer une liste de résultats de recherche que l'utilisateur doit parcourir plutôt que d'exécuter la tâche attendue. En revanche, ils reçoivent des réponses plus intuitives venant des solutions sédentaires. Les utilisateurs possédant les deux types de solutions (domestique et mobile) préférèrent la solution domestique car considérée comme plus fiable.
[1]	VOCH2	Population malvoyante	Six participants (6/14) ont partagé des expériences où le retour du système était verbeux, inutile, sans pertinence, ou inversement, manquait de précision pour satisfaire les requêtes formulées.
[1]	VOCH2 + VOCH8	Population malvoyante	Un participant (1/14) a souligné que les repères visuels n'étaient pas présentés efficacement à l'aide de moyens alternatifs. Par exemple, l'anneau lumineux de l'Amazon Echo clignote pour indiquer une notification de rappel et change de couleur pour indiquer la mise en sourdine du microphone. Le manque de considération pour les personnes malvoyantes est jugé frustrant.
[1]	VOCH13	Population malvoyante	Quatre participants (4/14) ont apprécié la commodité liée au contrôle de leur habitat à l'aide de leur assistant personnel. Les participants ont souligné qu'un meilleur sentiment d'indépendance pourrait être atteint en contrôlant plus de périphériques dans leur espace de vie.

Tableau A.6 – Préoccupations des populations à besoins spécifiques en matière d'interaction vocale

Étude	Heuristique	Groupe ciblé	Évaluation/Constatation
[1]	VOCH13	Population malvoyante	Les assistants personnels permettent aux utilisateurs d'effectuer diverses tâches plus rapidement et de réduire les efforts nécessaires pour interagir avec de nombreux services ou applications, dont certains sont inaccessibles aux lecteur d'écran, par rapport à l'utilisation d'un clavier tactile. Cependant, trois participants (3/14) ont mentionné la nécessité d'accéder vocalement à plus d'applications et de services.
[1]	VOCH4	Population malvoyante	Le processus d'identification des erreurs et de récupération de celles-ci s'est avéré être difficile selon les participants.
[1]	VOCH10	Population malvoyante	Les participants sont généralement disposés à accepter la menace de vol ou d'utilisation abusive de leurs données comme un compromis raisonnable pour avoir accès à des tâches autrement inaccessibles. Plusieurs participants ont indiqué que leurs données personnelles étaient déjà exposées de plusieurs autres manières (par exemple, par leur compte Gmail et iCloud) et donc ne s'inquiétaient pas d'autres violations.
[1]	VOCH10	Population malvoyante	La moitié des participants (7/14) ont préféré ne pas interagir verbalement fréquemment lors de trajets en transports en commun en raison du risque de divulgation d'informations d'identification et d'authentification, ou que des inconnus puissent entendre des éléments de leur vie privée.
[1]	VOCH10	Population malvoyante	Cinq participants (5/14) ont indiqué qu'une présentation plus rapide du contenu les aidait à préserver leur vie privée, en particulier en public ou en présence d'autres individus. D'après eux, une réponse très rapide du système est inintelligible pour la plupart des personnes sans trouble de la vue.
[1]	VOCH9	Population malvoyante	Pour surmonter les difficultés d'interaction par la parole dans les milieux bruyants, les participants ont en grande partie choisi d'utiliser un casque sur leurs appareils mobiles.
[1]	VOCH9	Population malvoyante	Neuf participants (9/14) ont décrit un malaise associée à l'utilisation de leur assistant personnel mobiles en public. Il peut être considéré comme étant socialement inapproprié d'interagir verbalement avec un appareil mobile à des moments particuliers selon les autres populations. En conséquence, les participants limitent leurs interactions en public.
[1]	VOCH9	Population malvoyante	Contrairement à la gêne ressentie dans les lieux publics, les participants ont généralement indiqué que leur domicile est considéré comme étant une «zone de confort» où ils n'ont aucune inquiétude à interagir vocalement.
[4]	VOCH7 + VOCH4	Population malvoyante	Deux participants (2/8) ont préféré utiliser le clavier plutôt que la parole en raison des difficultés d'édition. Tous les participants ont mentionné la rapidité comme principal avantage de la parole en tant que mode d'interaction.

Tableau A.7 – Préoccupations des populations à besoins spécifiques en matière d'interaction vocale

Étude	Heuristique	Groupe ciblé	Évaluation/Constatation
[4]	VOCH4	Population malvoyante	Bien que la parole soit préférée, tous les participants ont trouvé certains aspects de la saisie vocale frustrants. Tous les participants à l'exception d'un seul (7/8) ont cité l'édition comme source de frustration.
[4]	VOCH4	Population malvoyante	Lorsqu'ils interagissent avec la parole, les participants passent la plupart de leur temps à réviser et corriger les erreurs de reconnaissance. En moyenne, 80,3% (SD=10,2) du temps alloué à l'interaction est consacré à l'édition. En utilisant le clavier, en moyenne, les participants ont passé seulement 9,0% (SD=11,7) de leur temps à réviser et éditer leur texte.
[4]	VOCH4	Population malvoyante	L'édition incluait l'insertion, la suppression et le remplacement de caractères, de mots ou de groupes de mots. Seulement 15,6% de ces modifications ont été effectuées en utilisant la saisie vocale. Bien que la parole soit le mode de saisie le plus rapide, les participants ont préféré utiliser le clavier pour éditer le texte dicté.
[4]	VOCH4	Population malvoyante	Après avoir énoncé plusieurs phrases, la plupart des participants ont passé en revue leur texte en demandant au système de le lire mot par mot. Si un mot ne semblait pas correct, ils repassaient dessus lettre par lettre.
[25]	VOCH1	Population avec troubles physiques et population avec troubles de la parole	Certains participants se sont directement adressés aux appareils à contrôler tandis que d'autres se sont adressés à l'interface vocale, en lui demandant d'agir sur le l'environnement et de contrôler d'autres appareils. Pour ces derniers ordonner les appareils connectés séparément semblait peu naturel.
[25]	VOCH7 + VOCH1	Population avec troubles physiques et population avec troubles de la parole	Les participants ont utilisé une variété de styles pour s'adresser à l'interface vocale, allant d'une interaction purement «technique» au style injonctif à une interaction plus anthropomorphe et personnelle avec le système. Ce dernier style d'interaction se concentre sur l'interface vocale en tant que partenaire conversationnel. Le style injonctif met l'accent sur l'efficacité et la fiabilité.
[11]	VOCH12	Population âgée	37% des sondés (74/200) ne souhaitent pas d'apparence humaine pour l'interface car jugée trop effrayante. Ils sont plus enclins à l'accepter si le visage d'un proche est utilisé.
[11]	VOCH7	Population âgée	95% des sondés (190/200) ont déclaré qu'ils considéraient le système comme étant facile à utiliser en raison de son interface multimodale.
[11]	VOCH4	Population âgée	En cas d'incompréhension du système, la plupart des sondés ont répondu qu'ils répétaient la même commande jusqu'à ce que le système la comprenne.
[11]	VOCH3	Population âgée	5% des personnes interrogées (10/200) ont prétendu ne plus vouloir utiliser le système en cas d'erreurs répétées.
[11]	VOCH8	Population âgée	Parmi les inconvénients du système, les plus jeunes sondés ont indiqué avoir peur de perdre leurs compétences en laissant le système agir à leur place. Par exemple, ils ont affirmé qu'ils perdraient leurs capacités mémorielles si le système leur rappelait tout.

Tableau A.8 – Préoccupations des populations à besoins spécifiques en matière d'interaction vocale

Étude	Heuristique	Groupe ciblé	Évaluation/Constatation
[36]	VOCH1	Population âgée	Les participants ont exprimé des craintes relatives au manque de contrôle sur un système autonome. Au lieu de répondre à leurs attentes, le système exécute ce qu'on lui demande sans savoir si ce sont les mêmes choses.
[36]	VOCH9	Population âgée	Les participants sont désireux d'interagir avec un système proactif capable de fournir des informations pertinentes en fonction du contexte (p. ex., conseiller de prendre un parapluie en fonction de la météo sans avoir à demander le bulletin météorologique).
[36]	VOCH13	Population âgée	Les participants ont peur de perdre en diversité d'appareils contrôlable par la parole du fait que tous ne soient pas compatibles.
[36]	VOCH10	Population âgée	Les questions relatives au respect de l'intimité ont été soulevées mais ne sont pas pour autant une préoccupation importante pour les participants.
[85]	VOCH4	Population âgée	Tous les utilisateurs ont semblé être rapidement frustrés par les tentatives répétées de récupération après une erreur. La faute incombe à l'incapacité du système à motiver les erreurs de compréhension.
[85]	VOCH7	Population âgée	L'efficacité avec laquelle une tâche est exécutée n'est pas nécessairement corrélée à la satisfaction, en particulier dans le cas des personnes âgées pour lesquelles le temps n'est souvent pas un problème.
[85]	VOCH2	Population âgée	Un facteur clé important pour les personnes âgées est que les messages en sortie ou les instructions soient aussi courts que possible. Les messages devraient contenir au plus une information ou ne demander qu'une seule donnée à la fois.
[85]	VOCH11	Population âgée	Il est important pour les utilisateurs de maintenir leur confiance notamment grâce à des phrases de confirmation leur assurant une traduction correcte de leurs intentions.
[85]	VOCH6	Population âgée	Les adultes plus âgés sont sensibles au renforcement positif à mesure qu'ils apprennent à interagir avec un nouveau système.
[85]	VOCH5	Population âgée	L'apprentissage par l'exemple est particulièrement efficace pour les personnes âgées.
[77]	VOCH1	Population âgée	Tandis que les jeunes utilisateurs ont principalement énoncé des requêtes directement liées aux tâches à accomplir, les utilisateurs plus âgés ont souvent tenté d'interagir avec le système. Ils ont remercié le système de fournir des informations, ou ont agrémente leurs énoncés d'informations superflues et incompréhensibles par le système de dialogue. En particulier, les utilisateurs plus âgés ont souvent tenté de prendre l'initiative sur le dialogue, même si les dialogues étaient strictement à l'initiative du système.
[77]	VOCH3	Population âgée	La parole des utilisateurs plus âgés entraîne des taux d'erreur plus élevés que les utilisateurs plus jeunes du fait de facteurs à la fois acoustiques et linguistiques.

Tableau A.9 – Préoccupations des populations à besoins spécifiques en matière d'interaction vocale